UNIVERSITY OF CAPE TOWN

DEPARTMENT OF MATHEMATICAL STATISTICS

QUADRATIC PROGRAMMING AS AN EXTENSION OF LINEAR PROGRAMMING

by

R. L. Teixeira

A thesis prepared under the supervision of Professor C.G. Troskie in fulfilment of the requirements for the degree of Master of Science in Mathematical Statistics

TO VASCO, my father

A C K N O W L E D G E M E N T S

Many circumstances and persons have influenced the writing of this thesis, but it is not possible to give proper credit to all. However, I should like to mark my special debt to PROFESSOR C.G. TROSKIE, and make it quite clear that, if it were not for his guidance, encouragement, patience and above all youthful dynamism, this thesis would never have become the reality that it is today. To him alone I say: "Thank you".

I must also place on record the excellent co-operation and efficiency of MRS. M.I. COUSINS - responsible for the typing of this thesis. This task necessitated numerous hours of night work, and the loss of many weekends, and yet, I can never recall a word of complaint from this most diligent and charming lady.

My indebtness goes also to DR. R.I. BECKER for his careful preparation of Appendix III.

Cape Town, South Africa R.L.T.

October 1968.

P R E F A C E

In the past two decades Mathematical Programming has come to occupy a place of importance in Economic Studies and in Operations Research. Roughly speaking, Mathematical Programming is the analysis of problems of the type: "Find the maximum of a function, when the variables are subject to inequality and equality constraints". The term "Linear Programming" corresponds to the case where, the function to be maximized (the so called objective function) and the equality and inequality constraints are linear. The term "Non-Linear Programming" should then become self-defined.

With the introduction of Dantzig's Simplex Method, Linear Programming has become an everyday technique. The same, we regret to say, is not true for Nonlinear Programming because this subject is broader and much more difficult to unify than that of Linear Programming. In fact at present there does exist any unifying theory for Nonlinear Programming. However, we feel that research on this field is gathering tremendous momentum and that in the not to distant future Nonlinear Programming will become both a practical and fundamental tool in many spheres of Science.

One of the subject matters of Nonlinear Programming is what we came to call "Quadratic Programming". This name is restricted to the specific problem of maximizing or minimizing a quadratic objective function $f(X) = CX + X'DX$, where CX is a linear form and $X'DX$ a quadratic form, subject to linear constraints. Historically, Quadratic Programming was the first venture into the theory of Nonlinear Programming. This we feel, was due to the fact that (a) mathematically, quadratic programming is the natural first extension beyond the realm of linear programming and (b) in actual practice quadratic programming problems have the advantage of being "solvable" in a finite number of steps. In recent years, a large number of algorithms for quadratic programs (roughly, for the time being, we shall use this term to mean a quadratic programming problem) have been developed. Kunzi and Krelle (16) discuss seven of them, in addition to three versions of the application of gradient methods to quadratic programming. Since then a number of variants of Wolfe's Method (10) have been published. From this brief exposition the reader might conclude that, as far as the field of quadratic programming is concerned, he is faced with (i) a theory that is not unified and is therefore not important in its own right (ii) a variety of methods whose particular "idiosyncrasies" might leave him "flabbergasted" and undecided as to which is best to apply when dealing with a practical problem. Because of this and since Dantzig's Simplex Method for linear programming has proved to be an exceedingly convenient and efficient method for obtaining the extremum of a linear function subject to linear equalities and inequalities we hit upon the idea of presenting a "Simplex Approach" to the problem of finding the minimum maximum of a quadratic convex concave function subject to linear inequalities and equalities (the maximizing —— concavity and minimizing —— convexity restrictions must be imposed in order to prevent the existence of various local extrema). More specifically it is the purpose of this thesis to:

(i) Present a unified and simple treatment of the Theory of Concave (Convex) Quadratic Programming (in no way will mathematical rigour be sacrificed for simplicity).

(ii) Present a collection of "Simplicial Methods" for solving quadratic programming problems, which are but extensions of the Simplex Method for Linear Programming, whose "accuracy" and "convergence" make them completely self-sufficient for the solution of any type of concave (convex) quadratic programming problems. As to how we set about with this task the reader is referred to Chapter 0 where a more detailed discussion of the nature and structure of the present thesis is given.

R.L.T.

C O N T E N T S

PART II : SIMPLICIAL METHODS OF QUADRATIC PROGRAMMING.

Chapter 0

Introduction : Nature of Thesis

We have tried to make this thesis as self-contained as possible for any reader with a sound knowledge of the first eight Chapters of G. Hadley's book "Linear Programming" (see reference (14) in bibliography). In this Chapter we shall explain the organization of the thesis, give a brief review of the contents of each Chapter and explain what we hope to achieve.

ORGANIZATION:

The present thesis is divided into two parts. In Part I we present a unified treatment of the theory of Concave (Convex) Quadratic Programming. In Part II we present four computational methods which can rightfully claim to be an extension of the Simplex Method for Linear Programming to the solution of concave (convex) quadratic programs. For this reason we shall call these methods SIMPLEX METHODS for QUADRATIC PROGRAMMING or just SIMPLICIAL METHODS. PART I consists of three Chapters and PART II, four Chapters. Each Chapter is provided with an Appendix where additional proofs, explanations and references made in the Chapter are given. We find that this approach does not disrupt the continuity of the discussion being carried out in the main text and at the same time serves as a source of quick reference for the reader.

The style of presentation has been greatly influenced by the lectures of Professor C.G. Troskie at the University of Cape Town. Thus in Part I which is mainly concerned with theory, we have tried to eliminate most of the "padding" and trim the Chapters into the quick —— easy-to-follow pattern : Definition ——— Theorem ——— Remarks. Part II is a discussion of computational methods and is of necessity in essay form. Nevertheless we attempt to maintain a uniform structure by dividing each Chapter into five sections, namely:-

(1) Introduction (of the method).

(2) Theory of the Method.

(3) Description of the Method.

(4) Convergence of the Method.

(5) A "Flow" of the Method in Summary Form.

REVIEW OF CHAPTERS:

Chapter I. In this Chapter for the benefit of the readers not familiar with the "Theory of Convex and Concave Functions" and to make this thesis as self-contained as possible, we present those properties of "Convex Sets", "Convex Functions" and "Concave Functions" which are needed for our purposes.

Chapter II. In this Chapter we analyse in detail a special type of "concave program" (by a concave program we mean a problem of the form maximize a concave function subject to "convex constraints") namely "concave differentiable programs" and obtain a characterization for the optimal solutions of these programs. The main features of this Chapter are:

(i) An introduction of a hypothesis called "Hypothesis I" (which is motivated by the concept of "superconsistent differentiable programs"). Hypothesis I enables us to present a simpler and easier approach to the "Kuhn-Tucker Theory", avoiding the "constraint qualification" concept introduced by Kuhn and Tucker.(See reference

(7) in bibliography). This constraint qualification concept is somewhat difficult to grasp and although of fundamental importance in "more general" nonlinear programs we find that we can do away with it as far as the theory of "simplex" concave quadratic programming is concerned.

(ii) A presentation of the original Kuhn-Tucker Theorem (see reference (7) in bibliography), in a form which provides a unifying theoretical approach to the derivation of the theory of the "Simplicial Methods" presented in Part II.

Chapter III. In this Chapter we propose a classification for quadratic programs and apply the theory developed in Chapter II to obtain a characterization for the optimal solutions of quadratic programs. The important feature of this Chapter is the characterization of the optimal solution for a concave quadratic programming problem as a basic solution of a system of equations. Historically, it was Barankin and Dorfman (1) who first pointed out that, "if the linear Lagrangian conditions of optimality were combined with those of the original system, the optimum solution was a basic solution in the enlarged system with the property that only one of certain pairs of variables were in the basic set. Markowitz (8), on the other hand, proved that it was possible to modify the enlarged system and then parametrically generate a class of basic solutions with the above special property, which converged to the optimum in a finite number of steps". Both the articles of Markowitz and Barankin and Dorfman require rather involved arguments. Here we derive the "basic property" of the optimal solution to a concave quadratic programming problem in a simple but indirect way (see THM. IX.3 and THM. X.3). This is one instance where we hope to get away with murder without having to face the guillotine.

Chapter IV. In this Chapter a detailed presentation of Wolfe's Method (10) is given. Although the literature on this method is plentiful we feel that there are some "loose ends" which we try to patch up here as best as we can. These refer specially to the proof of the finite convergence of the method, to the usefulness of its "Long Form" in actual practice and to the application of its "Short Form" to quadratic programs where the objective function has no linear form.

Chapter V. In this Chapter a detailed presentation of the Method of Frank and Wolfe is given. The literature on this computational method is very sparse and in our opinion this is unjustified, for we find it quite useful in actual practice because of its high degree of finite convergence and simplicity (any computer program for the Simplex Method can be easily converted for the Method of Frank and Wolfe). Moreover, from the theoretical point of view it occupies a place of relevant importance in the present thesis, owing to the fact that the "basic property" of the optimal solution to the quadratic program (referred to in the sketch of Chapter III) is obtained as a "by product" of the finite convergence of the method. By this we mean that the "basic property" of the optimal solution is arrived at during the course of the proof of the finite convergence of the Method of Frank and Wolfe.

Chapter VI. In this Chapter we extend "Dantzig's Method". (See (5) in bibliography) to the case where the concave objective function consists of "mixed" quadratic and linear terms. Again, as in Chapter V, while discussing the finite convergence of this method we arrive at "another basic property" — namely "complementarity", of the optimal solution in question (see section (6.4) of Chapter VI).

Chapter VII. In this Chapter a very short presentation of Beale's Method is given. This should in no way indicate any ill feelings towards the method. As a matter of fact it remains our favourite, especially when actual numerical computations are involved. The reason for the brief treatment is due to the extremely lucid and exhaustive article on Beale's Method (by E.M.L. Beale himself) which appears in J. Adadie's book "Nonlinear Programming" (pp 143-172; see (11) in bibliography). We felt that any attempt at our own presentation of the method could in no way measure up to this excellent article.

(Note — the terminology THM.VI.3, say, is used to indicate that we are referring to THM. VI of Chapter III).

AIM of THESIS:

In this thesis we have tried to present an exposition of Concave (Convex) Quadratic Programming as an extension of Linear Programming. It is hoped that the thesis together with the article on "Beale's Method" mentioned above could serve as a text for a person acquainted with Linear Programming who seeks a systematic and simple extension of his knowledge to Concave Quadratic Programming.

PART I

QUADRATIC PROGRAMMING THEORY

Chapter I

"Convex Sets"

"Convex and Concave Functions"

(1.1) Convex Sets.

DEFN. Line Segment. The line segment joining two arbitrary but fixed points (vectors) X_1 and X_2 of the n-dimensional Euclidean space E_n is defined to be the set of points

$$\mathbb{Y} = \{X \mid X = \lambda X_2 + (1-\lambda)X_1 , \quad 0 \leq \lambda \leq 1\}$$

DEFN. Convex Set. A set $\mathbb{X}$ of points in E_n is said to be convex if the line segment joining any two points X_1 and X_2 of $\mathbb{X}$ is contained in $\mathbb{X}$, i.e., if for any two points X_1 and X_2 of $\mathbb{X}$ and for every $\lambda \in [0,1]$ the point $\lambda X_2 + (1-\lambda)X_1 \in \mathbb{X}$, then $\mathbb{X}$ is said to be convex.

LEMMA I If $\mathbb{A}$ and $\mathbb{B}$ are convex subsets of E_n, then $\mathbb{A} \cap \mathbb{B}$ is also a convex subset of E_n.

PROOF: Suppose X_1 and X_2 are any two points of $\mathbb{A} \cap \mathbb{B}$. Then $X_1 \in \mathbb{A}$, $X_1 \in \mathbb{B}$ and $X_2 \in \mathbb{A}$, $X_2 \in \mathbb{B}$. Hence

$$\lambda X_2 + (1-\lambda)X_1 \in \mathbb{A} \quad \text{for} \quad 0 \leq \lambda \leq 1$$

$$\lambda X_2 + (1-\lambda)X_1 \in \mathbb{B} \quad \text{for} \quad 0 \leq \lambda \leq 1 ,$$

since $\mathbb{A}$ and $\mathbb{B}$ are convex sets. Thus

$$\lambda X_2 + (1-\lambda)X_1 \in \mathbb{A} \cap \mathbb{B} \quad \text{for} \quad 0 \leq \lambda \leq 1.$$

Hence for any two points X_1 and X_2 of $\mathbb{A} \cap \mathbb{B}$, the line segment joining them belongs to $\mathbb{A} \cap \mathbb{B}$. Q.E.D.

REMARK: (i) The above theorem shows that convexity is preserved under set intersection.

(ii) It can now be easily shown that the intersection of any finite number of convex sets is convex. (Proof: Use induction. Result immediate.).

(1.2) Convex Functions.

DEFN. Convex Function. A function $f(X)$ defined on a convex subset $\mathbb{X}$ of E_n is said to be convex if for any two points X_1 and X_2 in $\mathbb{X}$ and for all λ, $0 \leq \lambda \leq 1$,

$$f(\lambda X_2 + (1-\lambda)X_1) \leq \lambda f(X_2) + (1-\lambda)f(X_1).$$

A convex function $f(X)$ defined on a convex subset $\mathbb{X}$ of E_n is said to be strictly convex if the strict inequality

$$f(\lambda X_2 + (1-\lambda)X_1) < \lambda f(X_2) + (1-\lambda)f(X_1)$$

is satisfied for all λ, $0 \leq \lambda \leq 1$, and for any two different points X_1 and X_2 belonging to $\mathbb{X}$.

LEMMA II If $g(X)$ is a convex function defined on a convex subset $\mathbb{X}$ of E_n and c is an arbitrary real number, then the set

$$\mathbb{Z} = \{X \mid g(X) \leq c \ , \quad X \in \mathbb{X}\}$$

is a convex subset of $\mathbb{X}$.

PROOF: Let X_1 and X_2 be any two points in $\mathbb{Z}$. Then $g(X_1) \leq c$ and $g(X_2) \leq c$. Now let $X = \lambda X_2 + (1-\lambda)X_1$ $0 \leq \lambda \leq 1$. Then $X \in \mathbb{X}$ since $\mathbb{X}$ is convex and therefore

$$g(X) = g(\lambda X_2 + (1-\lambda)X_1) \leq \lambda g(X_2) + (1-\lambda)g(X_1) \text{ since } g(X) \text{ is convex in } \mathbb{X}$$
$$\leq c \ .$$

Hence $X = \lambda X_2 + (1-\lambda)X_1 \in \mathbb{Z}$ for any $0 \leq \lambda \leq 1$.

Q.E.D.

NOTATION: If the function $f(X)$ defined on E_n is continuous over $\mathbb{X} \in E_n$ we write $f \in C$ over $\mathbb{X}$ (read f belongs to the class C of continuous functions over $\mathbb{X}$).

If $f(X)$ and all its first partial derivatives are continuous over $\mathbb{X}$ we write $f \in C'$ over $\mathbb{X}$.

If $f(X) \in C'$ over E_n we simply write $f \in C'$.

DEFN. Differentiable Function. If $f \in C'$ at X_o then $f(X)$ is said to be differentiable at X_o. If $f(X)$ is differentiable at every point of a set $\mathbb{X} \subset E_n$ we say that $f(X)$ is differentiable in $\mathbb{X}$.

DEFN; Gradient Vector. Let $f \in C'$ over $\mathbb{X}$, a subset of E_n. Then for every point X of $\mathbb{X}$ we can define an n-component row vector $\nabla f(X)$ by

$$\nabla f(X) = \left(\frac{\partial f(X)}{\partial x_1}, \ldots, \frac{\partial f(X)}{\partial x_n}\right) .$$

The vector $\nabla f(X)$ is called the gradient vector of $f(X)$.

THM. I If $f(X)$ is a convex function on E_n and if $f(X) \in C'$ then

$$f(X) - f(X_o) \geq \nabla f(X_o)(X-X_o)$$

for any fixed point $X_o \in E_n$ and for all $X \in E_n$.

PROOF: Let X_o be a fixed point in E_n. Then by the convexity of $f(X)$ we have that

$$f(\lambda X + (1-\lambda)X_o) \leq \lambda f(X) + (1-\lambda)f(X_o),$$

for any $X \in E_n$ and for all λ, $0 \leq \lambda \leq 1$. Hence

$$\frac{f(\lambda X+(1-\lambda)X_o)-f(X_o)}{\lambda} \leq f(X) - f(X_o) \quad \forall\ \lambda,\ 0 < \lambda \leq 1 \tag{1.1}$$

Expanding $f(X + (1-\lambda)X_o)$ by Taylor's theorem (see App. I.1) we obtain

$$f(X_o+\lambda(X-X_o)) = f(X_o) + \lambda\nabla f(X_o+\lambda\theta(X-X_o))(X-X_o) \quad \text{for some } \theta \in [0,1].$$

Thus relation (1.1) becomes

$$\nabla f(X_o+\lambda\theta(X-X_o))(X-X_o) \leq f(X) - f(X_o) .$$

On taking the limit as $\lambda \to 0$, we have that

$$\nabla f(X_o)(X-X_o) \leq f(X) - f(X_o) \quad \text{for any } X \in E_n.$$

Q.E.D.

THM. II If the functions $f_i(X)$ $i = 1,\ldots,N$, are convex on the same convex subset $\mathbb{X}$ of E_n, the linear combination

$$\sum_{i=1}^{N} c_i f_i(X)$$

is also a convex function on $\mathbb{X}$ when the coefficients c_i are non-negative constants.

PROOF: Let X_1 and X_2 be arbitrary points of $\mathbb{X}$. Then for all λ, $0 \leq \lambda \leq 1$

$$f_i(\lambda X_2 + (1-\lambda)X_1) \leq \lambda f_i(X_2) + (1-\lambda)f_i(X_1), \quad i = 1,\ldots,N,$$

because $f_i(X)$ is convex. Furthermore, since $c_i \geq 0$ $i = 1,\ldots,N$ we have that

$$c_i f_i(\lambda X_2 + (1-\lambda)X_1) \leq \lambda c_i f_i(X_2) + (1-\lambda)c_i f_i(X_1) \quad i = 1,\ldots,N.$$

Hence

$$\sum_{i=1}^{N} c_i f_i(\lambda X_2 + (1-\lambda)X_1) \leq \lambda \sum_{i=1}^{N} c_i f_i(X_2) + (1-\lambda) \sum_{i=1}^{N} c_i f(X_1)$$

Q.E.D.

THM. III Let $f(X) \in C'$ be a convex function on E_n. Then for any two given (fixed) points X_o and X_1 the function

$$\phi(\lambda) = \nabla f(X_o+\lambda(X_1-X_o))(X_1-X_o),$$

of the real variable λ, is a monotonically nondecreasing function.

PROOF: Let $\lambda_1 > \lambda_2$. Then by THM 1 we have that

$$f(X_o+\lambda_1(X_1-X_o))-f(X_o+\lambda_2(X_1-X_o)) \geq (\lambda_1-\lambda_2)\nabla f(X_o-\lambda_2(X_1-X_o))(X_1-X_o)$$

$$f(X_o+\lambda_2(X_1-X_o))-f(X_o+\lambda_1(X_1-X_o)) \geq (\lambda_2-\lambda_1)\nabla f(X_o+\lambda_1(X_1-X_o))(X_1-X_o)$$

Therefore result follows immediately on adding these two inequalities.

Q.E.D.

THM. IV Let $f(X) \in C'$ be a convex function on E_n. If for any two given (fixed) points X_o and X_1

$$\nabla f(X_o)(X_1-X_o) > 0$$

then

$$f(X_o+\lambda(X_1-X_o)) > f(X_o) \quad \text{for all} \quad \lambda > 0 .$$

PROOF: By THM I we have that

$$\begin{aligned} f(X_o+\lambda(X_1-X_o)) - f(X_o) &\geq \lambda\nabla f(X_o)(X_1-X_o) \\ &> \nabla f(X_o)(X_1-X_o) \quad \text{for all} \quad \lambda > 0 \\ &> 0 \quad \text{for all} \quad \lambda > 0 \end{aligned}$$

Q.E.D.

THM. V Let $f(X) \in C'$ be a convex function on E_n. If for any two given (fixed) points X_o and X_1

$$\nabla f(X_o)(X_1-X_o) < 0 ,$$

then there exists a $\lambda_o > 0$ such that

$$f(X_o+\lambda(X_1-X_o)) < f(X_o) \quad \text{for} \quad 0 < \lambda < \lambda_o .$$

PROOF: By THM I we have that

$$f(X_o)-f(X_o+\lambda(X_1-X_o)) \geq \nabla f(X_o+\lambda(X_1-X_o)) (-\lambda(X_1-X_o))$$

Hence

$$f(X_o+\lambda(X_1-X_o))-f(X_o) \leq \nabla f(X_o+\lambda(X_1-X_o))(X_1-X_o) \text{ for } 0 \leq \lambda \leq 1$$

Since the partial derivatives of $f(X)$ are continuous, we obtain, on taking one sided-limits

$$\lim_{\lambda\to 0^+} (f(X_o+\lambda(X_1-X_o)-f(X_o)) \leq \nabla f(X_o)(X_1-X_o) < 0 .$$

whence result follows immediately from the definition of a limit.

Q.E.D.

(1.3) Concave Functions.

DEFN. Concave function. A function $f(X)$ defined on a convex subset $\mathbb{X}$ of E_n, is said to be concave, if for any two points X_1 and X_2 in $\mathbb{X}$ and for all λ, $0 \leq \lambda \leq 1$,

$$f(\lambda X_2 + (1-\lambda)X_1) \geq \lambda f(X_2) + (1-\lambda)f(X_1) .$$

A concave function $f(X)$ defined on a convex subset $\mathbb{X}$ of E_n is said to be strictly concave if the strict inequality

$$f(\lambda X_2 + (1-\lambda)X_1) > \lambda f(X_2) + (1-\lambda)f(X_1)$$

is satisfied for all λ, $0 \leq \lambda \leq 1$, and for any two different points X_1 and X_2 belonging to $\mathbb{X}$.

REMARK: Note that the definition for concave functions is obtained from the definition for convex functions by simply reversing the direction of the defining inequality. This means that $f(X)$ is concave if, and only if, $-f(X)$ is convex. Hence each theorem of section (1.2) has an analog for concave functions. We give the statements of these theorems below. Their proofs follow exactly the same form as those given in section (1.2), except for the trivial modification mentioned.

THM. I* If $f(X)$ is a concave function on E_n and if $f(X) \in C'$ then

$$f(X) - f(X_o) \leq \nabla f(X_o)(X-X_o)$$

for any fixed point $X_o \in E_n$ and for all $X \in E_n$.

THM. II* If the functions $f_i(X)$ $i = 1,\dots,N$ are concave on the same convex subset $\mathbb{X}$ of E_n, the linear combination

$$\sum_{i=1}^{N} c_i f_i(X)$$

is also a concave function on $\mathbb{X}$, when the ooefficients c_i are non-negative constants.

THM.III* Let $f(X) \in C'$ be a concave function on E_n. Then for any two given (fixed) points X_o and X_1 the function

$$\phi(\lambda) = \nabla f(X_o+\lambda(X_1-X_o))(X_1-X_o) ,$$

of the real variable λ, is a monotonically non-increasing function.

THM. IV* Let $f(X) \in C'$ be a concave function on E_n. If for any two given (fixed) points X_o and X_1

$$\nabla f(X_o)(X_1-X_o) < 0$$

then

$$f(X_o+\lambda(X_1-X_o)) < f(X_o)$$

for all $\lambda > 0$.

THM. V* Let $f(X) \in C'$ be a concave function on E_n. If for any two given (fixed) points X_o and X_1

$$\nabla f(X_o)(X_1-X_o) > 0$$

then there exists a $\lambda_o > 0$ such that

$$f(X_o+\lambda(X_1-X_o) > 0$$

for all λ, $0 < \lambda < \lambda_o$.

(1.4) Maxima and Minima of Convex and Concave Functions.

DEFN. Global Maximum. Let $f(X)$ be a convex (concave) function defined on a convex subset $\mathbb{X}$ of E_n. Then $X_o \in \mathbb{X}$ is said to be a global maximum of $f(X)$ over $\mathbb{X}$ if

$$f(X) \leq f(X_o)$$

for all $X \in \mathbb{X}$.

DEFN. Global Minimum. Let $f(X)$ be a convex (concave) function defined on a convex subset $\mathbb{X}$ of E_n. Then $X_o \in \mathbb{X}$ is said to be a global minimum of $f(X)$ over $\mathbb{X}$ if

$$f(X_o) \leq f(X)$$

for all $X \in \mathbb{X}$.

DEFN. Local Maximum. Let $f(X)$ be a convex (concave) function defined on a convex subset $\mathbb{X}$ of E_n. Then $X_o \in \mathbb{X}$ is said to be a local maximum of $f(X)$ over $\mathbb{X}$ if

$$f(X) \leq f(X_o)$$

for each point X lying in both $\mathbb{X}$ and some sufficiently small neighbourhood of X_o, i.e., if there exists an $\varepsilon > 0$ such that

$$f(X) \leq f(X_o)$$

for all $X \in \mathbb{X} \cap \{X | \|X-X_o\| < \varepsilon\}$, then X_o is said to be a local maximum of $f(X)$ over $\mathbb{X}$.

DEFN. Local Minimum. Let $f(X)$ be a convex (concave) function defined on a convex subset $\mathbb{X}$ of E_n. Then $X_o \in \mathbb{X}$ is said to be a local minimum of $f(X)$ over $\mathbb{X}$ if

$$f(X_o) \leq f(X)$$

for each point X lying in both $\mathbb{X}$ and some sufficiently small neighbourhood of X_o, i.e., if there exists an $\varepsilon > 0$ such that

$$f(X_o) \leq f(X)$$

for all $X \in \mathbb{X} \cap \{X | \|X-X_o\| < \varepsilon\}$, then X_o is said to be a local minimum of $f(X)$ over $\mathbb{X}$.

REMARK: If $X_o \in \mathbb{X}$ is a local maximum of $f(X)$ over $\mathbb{X}$ then we say that $f(X)$ takes on a local maximum at X_o.

If $X_o \in \mathbb{X}$ is a global maximum of $f(X)$ over $\mathbb{X}$ then the unique real number $f(X_o)$ is called the global maximum (supremum) of $f(X)$ over $\mathbb{X}$ and we say that $f(X)$ takes on or attains the (its) global maximum (supremum) at X_o.

<u>THM. VI</u> Let $f(X)$ be a convex function defined on a convex subset $\mathbb{X}$ of E_n. Then any local minimum, X_o, of $f(X)$ over $\mathbb{X}$ is also a global minimum of $f(X)$ over $\mathbb{X}$.

PROOF: Suppose there exists an $X_* \in \mathbb{X}$ such that $f(X_o) > f(X_*)$. Then, since $f(X)$ is convex

$$f(\lambda X_* + (1-\lambda)X_o) \leq \lambda f(X_*) + (1-\lambda)f(X_o)$$

for all λ, $0 \leq \lambda \leq 1$.

$$\therefore f(\lambda X_* + (1-\lambda)X_o) < \lambda f(X_o) + (1-\lambda)f(X_o) = f(X_o) \qquad (1.2)$$

for all λ, $0 < \lambda < 1$, since $f(X_*) < f(X_o)$.

Now since X_o is a local minimum there exists an $\varepsilon > 0$, such that

$$f(X_o) \leq f(X)$$

for all $X \in \mathbb{X}$ $\{X \mid \|X-X_o\| < \varepsilon\} = N_\varepsilon$. Let λ_1 be such that $0 < \lambda_1 < \min\{1, \frac{\varepsilon}{\|X_*-X_o\|}\}$. Then the point $X = \lambda_1 X_* + (1-\lambda_1)X_o$ belongs to N_ε. But by (1.2)

$$f(\lambda_1 X_* + (1-\lambda)X_o) < f(X_o).$$

This contradicts the fact that $X_o \in \mathbb{X}$ is a local minimum of $f(X)$ over $\mathbb{X}$. Hence if X_o is a local minimum of $f(X)$ over $\mathbb{X}$ there cannot exist a point $X_* \in \mathbb{X}$ such that $f(X_o) > f(X_*)$. This implies that $f(X) \geq f(X_o)$ for all $X \in \mathbb{X}$, i.e., that X_o is a global minimum of $f(X)$ over $\mathbb{X}$.

Q.E.D.

<u>THM.VII</u> Let $f(X)$ be a convex function defined on a convex subset $\mathbb{X}$ of E_n. Then the set of points for which $f(X)$ takes on its global minimum is convex.

PROOF: If $f(X)$ is a convex function without a global minimum over $\mathbb{X}$ then the set of points for which $f(X)$ takes on its global minimum is empty and hence convex (see App. I.2). Assume then that the global minimum is taken on at a point $X_1 \in \mathbb{X}$. Let $\mathbb{B} = \{X | X \in \mathbb{X}, f(X) = f(X_1)\}$. If $\mathbb{B}$ has only one point (namely X_1) then clearly $\mathbb{B}$ is convex. Suppose then that there exists another point $X_2 \in \mathbb{B}$ such that $X_2 \neq X_1$.

Now

$$f(\lambda X_2 + (1-\lambda)X_1) \leq \lambda f(X_2) + (1-\lambda)f(X_1) = f(X_1) \text{ for all } \lambda,\ 0 \leq \lambda \leq 1 \quad (1.3)$$

But $\lambda X_2 + (1-\lambda)X_1 \in \mathbb{X}$ since $\mathbb{X}$ is convex and $\mathbb{B} \subseteq \mathbb{X}$ and so since $f(X) \geq f(X_1)$ for all $X \in \mathbb{X}$ it follows from (1.3) that

$$f(\lambda X_2 + (1-\lambda)X_1) = f(X_1) \text{ for all } \lambda,\ 0 \leq \lambda \leq 1 .$$

Hence $\lambda X_2 + (1-\lambda)X_1 \in \mathbb{B}$ for all λ, $0 \leq \lambda \leq 1$.

Q.E.D.

REMARK: Note that an immediate implication of THM. VII is that if the global minimum of a convex function, defined on a convex set $\mathbb{X} \subseteq E_n$, is taken on at two different points of $\mathbb{X}$, it is taken on at an infinite number of points, since then any point lying on the line segment joining the two points in question will be a global minimum.

<u>THM.VIII</u> Let $f(X)$ be a <u>strictly</u> convex function defined on a convex subset $\mathbb{X}$ of E_n. Then $f(X)$ can only take on its global minimum over $\mathbb{X}$ at a unique <u>point</u> of $\mathbb{X}$.

PROOF: Let $X_o \in \mathbb{X}$ be such that

$$f(X) \geq f(X_o)$$

for all $X \in \mathbb{X}$, and suppose there exists another point $X_1 \in \mathbb{X}$, $X_1 \neq X_2$ such that

$$f(X_1) = f(X_o) .$$

Now since $f(X)$ is strictly convex in $\mathbb{X}$ we have that

$$f(\lambda X_1 + (1-\lambda)X_o) < \lambda f(X_1) + (1-\lambda)f(X_o) \quad \forall\ \lambda,\ 0 \leq \lambda \leq 1 .$$

In particular for the point $X_2 = \frac{1}{2}X_1 + \frac{1}{2}X_o \in \mathbb{X}$, we have that

$$f(X_2) < \tfrac{1}{2}f(X_1) + \tfrac{1}{2}f(X_o) = f(X_o) .$$

But this is a contradiction since $f(X)$ takes on its global minimum over $\mathbb{X}$ at X_o.

Q.E.D.

REMARK: Note that THM.VIII immediately implies that if a convex function $f(X)$ defined on a convex set $\mathbb{X}$ of E_n takes on the global minimum over $\mathbb{X}$ at two different points of $\mathbb{X}$ then $f(X)$ cannot be a strictly convex function.

THM. IX Let $f(X)$ be a convex function defined on a convex subset $\mathbb{X}$ of E_n. Then $f(X)$ takes on its global minimum over $\mathbb{X}$ at every point in $\mathbb{X}$ which satisfies $\nabla f(X) = 0$.

PROOF: Let $X_o \in \mathbb{X}$ be such that $\nabla f(X_o)$. Then by THM. I

$$f(X) - f(X_o) \geq \nabla f(X_o)(X-X_o)$$

for all $X \in E_n$ and in particular for all $X \in \mathbb{X}$.

Hence

$$f(X) \geq f(X_o)$$

for all $X \in \mathbb{X}$.

Q.E.D.

Throughout this section we have neglected the problem of maximizing convex functions over convex sets. Let us now turn our attention to this subject.

THM. X If a convex function $f(X)$ has a global maximizing point in the interior of its convex domain $\mathbb{X}$, then f is constant on $\mathbb{X}$.

PROOF: (See next page).

PROOF: Let $f(X) \leq f(X_1)$ for all $X \in \mathbb{X}$ and suppose that X_1 is an interior point of $\mathbb{X}$. Now by the definition of interior point there exists an ε-nbhood $N_\varepsilon = \{X \mid \|X-X_1\| < \varepsilon\}$ around X_1 such that $N_\varepsilon \subseteq \mathbb{X}$.

Let X be an arbitrary point of $\mathbb{X}$ and consider the points

$$X(t) = \frac{\|X_1 - tX\|}{(1-t)} \quad 0 \leq t \leq 1$$

Now since $\lim_{t \to 0} \|X(t) - X_1\| = 0$ it follows that for any ε_1, $0 < \varepsilon_1 < \varepsilon$, there exists a $0 < \delta < 1$ such that

$$\|X(t) - X_1\| < \varepsilon_1 < \varepsilon \quad \text{for all} \quad 0 < |t| < \delta < 1$$

Let λ be a number such that $0 < \lambda < \delta$. Then $X(\lambda) \in N_\varepsilon$ $\mathbb{X}$.

Denoting $X(\lambda)$ by X_2 we conclude that

$$X_1 = (1-\lambda)X_2 + \lambda X$$

for some point X_2 in $\mathbb{X}$ and some λ in $(0,1)$. Hence

$$f(X_1) \leq (1-\lambda)f(X_2) + \lambda f(X)$$

by the convexity of f. Now

$$f(X_2) \leq f(X_1) \quad \text{and} \quad f(X) \leq f(X_1), \tag{1.4}$$

therefore if $f(X) < f(X_1)$ relations (1.4) show that

$$f(X_1) < (1-\lambda)f(X_2) + \lambda f(X) = f(X_1)$$

which is a contradiction. Thus $f(X) = f(X_1)$ for all $X \in \mathbb{X}$.

As immediate consequences of THM. X we have:

(i) Every global maximizing point X_* lies on the boundary of the convex domain $\mathbb{X}$ over which a nonconstant convex function $f(X)$ is being maximized. Moreover, no such point can have $\nabla f(X_*) = 0$ for then X_* would be a global minimizing point for $f(X)$ (THM.IX), and $f(X)$ is a constant when a point on its domain $\mathbb{X}$ is both global maximum and a global minimum over $\mathbb{X}$.

(ii) Every local maximizing point $X_{\dagger}$ lies on the boundary of the convex domain $\mathbb{X}$ over which a nonconstant convex function $f(X)$ is being maximized, since if it did lie on the interior of $\mathbb{X}$ then $\nabla f(X_{\dagger}) = 0$ which would imply $f(X)$ = constant over $\mathbb{X}$, a contradiction.

So if the convex function $f(X)$ defined over a convex set $\mathbb{X}$ is not constant on $\mathbb{X}$ and has a global maximum X_*, say, on $\mathbb{X}$, then X_* lies on the boundary of $\mathbb{X}$. As a matter of fact it can be proved that :-
if $f(X)$ is a convex function defined over a closed convex set $\mathbb{X}$ bounded from below and if $f(X)$ takes on the global maximum at some point of $\mathbb{X}$, then the global maximum of $f(X)$ over $\mathbb{X}$ will be taken on at one or more extreme points of $\mathbb{X}$. (See App.I.3).
Since quadratic programming is not concerned with the problem of maximizing convex functions, we shall not pursue this subject further. Nevertheless quadratic programming is concerned with minimizing convex quadratic functions (or equivalently with maximizing concave quadratic functions) over convex sets. This subject will be analysed in detail in Chapter II.
Finally note that since if $f(X)$ is concave $-f(X)$ is convex and since maximum $f(X) = -$ minimum $|-f(X)|$ over $\mathbb{X}$, each theorem proved in this section, has an analog for concave functions:

(see following page)

THM. VI* Let $f(X)$ be a concave function defined on a convex subset $\mathbb{X}$ of E_n. Then any local maximum of $f(X)$ over $\mathbb{X}$ is also a global maximum of $f(X)$ over $\mathbb{X}$.

THM. VII* Let $f(X)$ be a concave function defined on a convex subset $\mathbb{X}$ of E_n. Then the set of points for which $f(X)$ takes on its global maximum is convex.

THM. VIII* Let $f(X)$ be a strictly concave function defined on a convex subset $\mathbb{X}$ of E_n. Then $f(X)$ can only take on its global maximum over $\mathbb{X}$, at a unique point of $\mathbb{X}$.

THM. (IX*) Let $f(X)$ be a concave function defined on a convex subset $\mathbb{X}$ of E_n. Then $f(X)$ takes on its global maximum over $\mathbb{X}$ at every point in $\mathbb{X}$ which satisfies $\nabla f(X) = 0$.

THM. X* If a concave function $f(X)$ has a global minimizing point in the interior of its convex domain $\mathbb{X}$, then $f(X)$ is constant on X.

Chapter II

"Convex and Concave Differentiable Programs"

(2.1) General Facts

DEFN. Differentiable Program. If $f(X) \in C'$ and if $g_i(X)$, $i = 1,\ldots,m$, is a convex function on E_n which belongs to C', then a differentiable program is defined to be a problem of the form:

$$\text{maximise or minimise } f(X) \tag{2.1}$$

subject to

$$\left.\begin{array}{ll} g_i(X) \leq 0 & i = 1,\ldots,m \\ X \geq 0 & \end{array}\right] \tag{2.2}$$

REMARK: In the sequel we shall use the abbreviation "D.P." to denote "Differentiable Program".

DEFN. Objective function. The function $f(X)$ in (2.1) is called the objective function.

DEFN. Feasible Solution. A vector (point) X which satisfies the constraints in (2.2) is called a feasible solution (point).

DEFN. Feasible Domain. The set $R_X = \{X | X \geq 0,\ g_i(X) \leq 0 \quad i = 1,\ldots,m\}$, of all feasible points is called the feasible domain. With R_X we can associate three important concepts:

(a) Boundary Surfaces. The hypersurfaces given by $g_i(X) = 0$ and $x_j = 0$ are called boundary surfaces of the feasible domain R_X.

(b) Boundary Point. A feasible point X_* is called a boundary point if it lies on at least one of the boundary surfaces.

(c) Interior Point. A feasible point X_* is called an interior point if $g_i(X_*) < 0$ for all i, and $X_* > 0$.

REMARK: Note that the feasible domain R_X is a closed convex set (see Appendix II.1)

DEFN. Optimal Solution. A feasible point X_*, for which $f(X)$ attains its maximum (minimum) over R_X, is called an optimal solution or simply "solution" to the D.P. . That is, if

$$f(X_*) \geq f(X) \quad \text{for all} \quad X \in R_X$$

then X_* is said to be an optimal solution to the D.P.

DEFN. Consistent D.P. A D.P. is said to be consistent (feasible) if its feasible domain R_X is not empty.

DEFN. Superconsistent D.P. A D.P. is said to be superconsistent if it is consistent and if in addition there exists a feasible point X_* such that

$$g_i(X_*) < 0 \quad \underline{\text{for all } i} \; .$$

DEFN. Solvable D.P. A feasible D.P. is solvable if the objective function is bounded over R_X and attains its maximum (minimum) over R_X.

DEFN. Concave Differentiable Program. A concave differentiable program is a differentiable program in which we specifically want to maximize an objective function $f(X)$ which is concave over E_n.

DEFN. Convex Differentiable Program. A convex differentiable program is a differentiable program in which we specifically want to minimize an objective function $f(X)$ which is convex over E_n.

REMARK: Note that it is the concavity or convexity of the objective function that makes the D.P. concave or convex.

THM. I Let $X_o = [x_1^o, \ldots, x_n^o]$ be a solution (optimal) to the concave differentiable program.

(2.3)
$$\left.\begin{array}{ll} \max f(X) & \\ \text{subject to} & \\ g_i(X) \leq 0 & i = 1,\ldots,m \\ X \geq 0 & \end{array}\right| \quad (2.3)$$

Then X_o is an optimal solution to the simplified differentiable program.

(2.4)
$$\left.\begin{array}{ll} \max f(X) & \\ \text{subject to} & \\ g_i(X) \leq 0 & \text{for } i \in I = \{i | g_i(X_o) = 0\} \\ x_j \geq 0 & \text{for } j \in J = \{j | x_j^o = 0\} \end{array}\right| \quad (2.4)$$

PROOF: Assume that X_o is not an optimal solution to (2.4). Then there exists a point X with

$$x_j \geq 0 \quad \text{for } j \in J$$
$$g_i(X) \leq 0 \quad \text{for } i \in I$$

such that $f(X) > f(X_o)$.

Now $x_j^o + \lambda(x_j - x_j^o) \geq 0$ for $j \in J$ and for $\lambda \geq 0$

$x_j^o + \lambda(x_j - x_j^o) \geq 0$ for $j \notin J$ and for sufficiently small $\lambda > 0$

Also

$$\begin{aligned} g_i(X_o + \lambda(X - X_o)) &\leq g_i(X_o) + \lambda(g_i(X) - g_i(X_o)) && \text{for } 0 \leq \lambda \leq 1 \\ &\leq \lambda g_i(X) && \text{for } i \in I \text{ and for } 0 \leq \lambda \leq 1 \\ &\leq 0 && \text{for } i \in I \text{ and for } 0 \leq \lambda \leq 1 \end{aligned}$$

$$g_i(X_o+\lambda(X-X_o)) \leq g_i(X_o) + \lambda(g_i(X)-g_i(X_o)) \quad \text{for} \quad 0 \leq \lambda \leq 1$$

$$\leq 0 \quad \text{for} \quad i \notin I \quad \text{and for sufficiently small } \lambda>0$$

Hence for sufficiently small $\lambda > 0$

$$g_i(X_o+\lambda(X-X_o)) \leq 0 \qquad \text{for all} \quad i$$

$$x_j^o +\lambda(x_j-x_j^o) \geq 0 \qquad \text{for all} \quad j$$

which implies that $X_o+\lambda(X-X_o)$ is a feasible point of the D.P. (2.3).

But $f(x)$ is concave

$$\therefore \quad f(X_o+\lambda(X-X_o)) \geq f(X_o) + \lambda(f(X)-f(X_o))$$

$$> f(X_o) \quad \text{since} \quad \lambda > 0, \quad f(x) > f(X_o) .$$

This contradicts the fact that X_o is an optimal solution of the D.P. (2.3)

Q.E.D.

REMARK: Note that in the proof of THM I we don't use the fact that $f \in C'$ and $g_i(X) \in C'$, $i = 1,\dots,m$. (See App.III.2)

(2.2) Saddle Points.

DEFN. Saddle Point. A function $F(X,\lambda)$, X being an n-component column vector, λ an m component column vector, defined in a subset (region) R of E_{n+m} is said to have a saddle point (global) at $[X_o,\lambda_o]$ in the region R if,

$$F(X,\lambda_o) \leq F(X_o,\lambda_o) \leq F(X_o,\lambda)$$

holds, for all $[X,\lambda] \in R$.

DEFN. Gradient of $F(X,\lambda)$. If $F(X,\lambda) \in C'$ over E_{n+m} we define the gradient of $F(X,\lambda)$ with respect to X evaluated at $[X_o,\lambda_o]$ to be

$$\nabla_X F(X_o,\lambda_o) = \left(\frac{\partial}{\partial x_1} F(X_o,\lambda_o),\dots, \frac{\partial}{\partial x_n} F(X_o,\lambda_o)\right)$$

$x_k^o > 0$ there exists an $\varepsilon > 0$ such that the point $[X_0 + he_k, \lambda_o]$ belongs to R^+ for all h, $0 < [h] < \varepsilon$. Therefore, since by assumption

$$F(X,\lambda_o) \leq F(X_o,\lambda_o) \quad \text{for all} \quad [X_1\lambda] \in R^+ ,$$

we have that

$$F(X_o + he_k) - F(X_o,\lambda_o) \quad \text{for all} \quad h, \quad 0 < [h] < \varepsilon.$$

Hence

$$\frac{F(X_o+he_k)-F(X_o,\lambda_o)}{h} \leq 0 \quad \text{for} \quad -\varepsilon < h < 0 \tag{2.7}$$

$$\frac{F(X_o+he_k)-F(X_o,\lambda_o)}{h} \geq 0 \quad \text{for} \quad 0 < h < \varepsilon \tag{2.8}$$

Since $F(X,\lambda) \in C'$ over E_{n+m} on taking limits as $h \to 0$ we conclude that

$$\frac{\partial}{\partial x_k} F(X_o,\lambda_o) \leq 0 \qquad \text{from} \quad (2.7)$$

$$\frac{\partial}{\partial x_k} F(X_o,\lambda_o) \geq 0 \qquad \text{from} \quad (2.8)$$

Thus $\quad \frac{\partial}{\partial x_k} F(X_o,\lambda_o) = 0 .$

Hence we have proved that if $[X_o,\lambda_o]$ is a global saddle point of $F(X,\lambda)$ in R^+ then for any component x_k^o of X_o we have

$$\frac{\partial}{\partial x_k} F(X_o,\lambda_o) = 0 \quad \text{if} \quad x_k^o > 0$$

$$\text{or} \quad \frac{\partial}{\partial x_k} F(X_o,\lambda_o) \leq 0 \quad \text{if} \quad x^o{}_k = 0$$

Therefore (2.5) follows immediately.

The proof of (2.6) follows exactly the same lines.

Q.E.D.

<u>THM.III</u> If $F(X,\lambda) \in C'$ over E_{n+m}, then a sufficient condition for $[X_o,\lambda_o] \in R^+ = \{[X,\lambda] \,|\, X \geq 0,\ \lambda \geq 0\}$ to be a saddle point (global) of $F(X,\lambda)$ in the region R^+ is that

$$\nabla_X F(X_o,\lambda_o) \leq 0 \quad \nabla_X F(X_o,\lambda_o)X_o = 0 \tag{2.9}$$

$$\nabla_\lambda F(X_o,\lambda_o) \geq 0 \quad \nabla_\lambda F(X_o,\lambda_o)\lambda_o = 0 \tag{2.10}$$

$$F(X,\lambda_o) \leq F(X_o,\lambda_o) + \nabla_X F(X_o,\lambda_o)(X-X_o) \quad \forall\ X \geq 0 \tag{2.11}$$

$$F(X_o,\lambda) \geq F(X_o,\lambda_o) + \nabla_\lambda F(X_o,\lambda_o)(\lambda-\lambda_o) \quad \forall\ \lambda \geq 0 \tag{2.12}$$

PROOF: $\nabla_X F(X_o,\lambda_o)(X-X_o) = \nabla_X F\ (X_o,\lambda_o)X$ since $\nabla_X F(X_o,\lambda_o)X_o = 0$

≤ 0 for all $X \geq 0$ since $\nabla_X F(X_o,\lambda_o) \leq 0$

Therefore from (2.11) it follows that

$$F(X,\lambda_o) \leq F(X_o,\lambda_o) \quad \forall\ X \geq 0$$

Similarly, using relations (2.10) and (2.12) we obtain

$$F(X_o,\lambda_o) \leq F(X_o,\lambda) \quad \forall\ \lambda \geq 0 .$$

Therefore for any given $[X,\lambda] \in R^+$ we have that

$$F(X,\lambda_o) \leq F(X_o,\lambda_o) \leq F(X_o,\lambda) .$$

Hence $[X_o,\lambda_o]$ is a global saddle point of $F(X,\lambda)$ in R^+. Q.E.D.

(2.3) Characterisation of an Optimal Solution of a Concave Differentiable Program.

In this section we consider two closely related problems namely, Problem I and Problem II, and derive very important relationships between the two. We start by introducing Problem I:

Problem I. Suppose that $f(X)$ is a concave function on E_n and that $f(X) \in C'$ over E_n. Suppose further that $g_i(X)$ is a convex function on E_n and that $g_i(X) \in C'$ for $i = 1,\ldots,m$. Find a point X_o which

maximizes $f(X)$

subject to

$$g_i(X) \leq 0 \qquad i = 1,\ldots,m$$

$$X \geq 0$$

DEFN. Lagrange Function. Given Problem I the function $F(X,\lambda)$ defined by

$$F(X,\lambda) = f(X) - \lambda' G(X)$$

where: $\lambda = [\lambda_1,\dots\dots\dots,\lambda_m]$

$$G(X) = |g_1(X),\dots,g_m(X)|,$$

is called the Lagrange function of Problem I.

Problem II. Given Problem I, find a point $[X_o,\lambda_o]$, such that $[X_o,\lambda_o]$ is a saddle point (global), of its Lagrange function $F(X,\lambda) = f(x)-\lambda' G(X)$, in the region $R^+ = \{[X,\lambda] \mid X \geq 0, \lambda \geq 0\}$.

THM. IV Suppose that the concave differentiable program given in Problem I is superconsistent and suppose that X_o is an optimal solution to Problem I. Then

$$\nabla f(X_o)(X-X_o) \leq 0$$

for all X satisfying the inequalities

$$\begin{cases} \nabla g_i(X_o)(X-X_o) \leq 0 & i \in I = \{i \mid g_i(X_o) = 0\} \\ X \geq 0 & \end{cases}$$

PROOF: Since $g_i(X)$ $i = 1,\dots,m$, is convex and differentiable in E_n, it follows from THM. I.1 that

$$g_i(X_o) + \nabla g_i(X_o)[X-X_o] \leq g_i(X) \qquad i = 1,\dots,m$$

$$\therefore \quad \nabla g_i(X_o)[X-X_o] \leq g_i(X) \quad \text{for } i \in I$$

Hence

$$\nabla g_i(X_o)[X_*-X_o] < 0 \qquad i \in I \qquad (2.13)$$

for some vector X_*, because the concave differentiable program in Problem I is superconsistent. Now let $X \geq 0$ be an arbitrary vector for which

$$\nabla g_i(X_o)(X-X_o) \leq 0 \qquad i \in I \qquad (2.14)$$

Multiplying inequality (2.13) by θ and inequality (2.14) by $(1-\theta)$, where $0 < \theta \leq 1$, and combining the resulting inequalities we obtain

$$(1-\theta)\nabla g_i(X_o)(X-X_o) + \theta\nabla g_i(X_o)(X_*-X_o) < 0 \quad i \in I, \quad 0 < \theta \leq 1,$$

which is equivalent to

$$\nabla g_i(X_o)\ ((1-\theta)X + \theta X_* - X_o) < 0 \qquad i \in I, \quad 0 < \theta \leq 1, \tag{2.15}$$

If we put

$$Z = (1-\theta)X + \theta X_* \qquad 0 < \theta \leq 1\ , \tag{2.16}$$

inequality (2.15) can be written in the more compact form

$$\nabla g_i(X_o)(Z-X_o) < 0 \qquad i \in I \tag{2.17}$$

Now by THM. I.1

$$g_i(X_o) - g_i(X_o+t(Z-X_o)) \geq (\nabla g_i(X_o+t(Z-X_o))(-t(Z-X_o))$$

$$\therefore\ g_i(X_o+t(Z-X_o)) - g_i(X_o) \leq t(\nabla g_i(X_o+t(Z-X_o))(Z-X_o), \tag{2.18}$$

so since $g_i(X)$ and its first partial derivatives are assumed to be continuous for all i, we have that for sufficiently small and positive t

$$g_i(X_o+t(Z-X_o)) < 0 \quad i \notin I \quad (\text{because } g_i(X_o) < 0,\ i \notin I)$$

$$g_i(X_o+t(Z-X_o)) < 0 \quad i \in I \quad (\text{because of relation (2.17)})\ .$$

Note: It is to conclude this very last statement that we need relation (2.17) as a strict inequality. For if we had $\nabla g_i(X_o)(Z-X_o) = 0$, $\nabla g_i(X_o+t(Z-X_o))(Z-X_o)$ could be strictly greater than 0 for sufficiently small and positive t. Also recall that we were able to obtain relation (2.17) as a strict inequality because the concave differentiable program in question is assumed to be superconsistent.

Whence for sufficiently small and positive t

$$g_i(X_o+t(Z-X_o)) < 0 \qquad i = 1,\ldots,m \tag{2.19}$$

Also for sufficiently small and positive t

$$X_o + t(Z-X_o) \geq 0 \tag{2.20}$$

since, by (2.16), $Z = (1-\theta)X + \theta X_*$ $0 < \theta \leq 1$, and $X \geq 0$, $X_* \geq 0$.
From (2.19) and (2.20) we conclude that for sufficiently small and positive t the point $X_o + t(Z-X_o)$ is a feasible solution to Problem I and therefore

$$f(X_o+t(Z-X_o)) - f(X_o) \leq 0 \tag{2.21}$$

for sufficiently small and positive t, because X_o maximizes $f(X)$.
Finally by THM. I^*.1

$$f(X_o+t(Z-X_o)) - f(X_o) \geq t\nabla f(X_o+t(Z-X_o))(Z-X_o) .$$

Hence, it follows from relation (2.21) that

$$\nabla f(X_o+t(Z-X_o))(Z-X_o) \leq 0 ,$$

for sufficiently small and positive t.

Letting t approach zero, we obtain

$$\nabla f(X_o)(Z-X_o) \leq 0$$

But by (2.16) this means that

$$\nabla f(X_o)((1-\theta)X + \theta X_* - X_o) < 0 \quad 0 < \theta \leq 1,$$

which reduces to

$$\nabla f(X_o)(X-X_o) \leq 0$$

on letting θ tend to zero.

Thus we have proved that

$$\nabla f(X_o)(X-X_o) \leq 0$$

when $X \geq 0$ satisfies the inequalities

$$\nabla g_i(X_o)(X-X_o) \leq 0 \qquad i \in I .$$

Q.E.D.

REMARK: THM. IV tells us that if we have a concave differentiable program <u>which is superconsistent</u> then for any optimal solution X_o (if any) of this program we must have that

$$\left.\begin{array}{l}\nabla f(X_o)(X-X_o) \leq 0 \\ \text{for all } X \text{ satisfying} \\ \qquad \nabla g_i(X_o)(X-X_o) \leq 0 \quad i \in I = \{i|g_i(X_o) = 0\} \\ \qquad\qquad X \geq 0\end{array}\right| \text{———}(2.22)$$

If we have a concave differentiable program which is not superconsistent and which has an optimal solution then assertion (2.22) may or may not be true. As it will be seen in the sequel assertion (2.22) plays a most important role, not only in theory of quadratic programming, but also in the general theory of non-linear programming. Thus, we find it convenient to introduce the following hypothesis:

HYPOTHESIS I. If for any optimal solution X_o (provided there exists one of course) of a concave differentiable program we have that

$$\nabla f(X_o)(X-X_o) \leq 0$$

for all X satisfying

$$\nabla g_i(X_o)(X-X_o) \leq 0 \quad i \in I = \{i|g_i(X_o) = 0\}$$

$$X \geq 0 ,$$

then we say that the concave differentiable program satisfies Hypothesis I.

REMARK: From THM. IV it is obvious that a superconsistent concave differentiable program satisfies Hypothesis I. In Section (2.4) we shall prove that any concave "quadratic" program also satisfies Hypothesis I.

THM. V If the concave differentiable program given in Problem I satisfies Hypothesis I and if X_o is an optimal solution to Problem I, then there exists an m-component column vector $\lambda_o \geq 0$ and an n-component row vector $\mu_o \geq 0$ such that

(i) $\nabla f(X_o) = \lambda_o' \nabla G(X_o) - \mu_o$ (where $\nabla G(X) = [\nabla g_1(X), \ldots, \nabla g_m(X)]$)

(ii) $\mu_o X_o = 0$

(iii) $\lambda_o' G(X_o) = 0$.

PROOF: Since X_o is an optimal solution to Problem I and the concave D.P. given in Problem I satisfies Hypothesis I we have that

$$\nabla f(X_o)(X - X_o) \leq 0$$

for all X satisfying the inequalities.

$$\begin{cases} \nabla g_i(X_o)(X - X_o) \leq 0 & i \in I = \{i \mid g_i(X_o) = 0\} \\ X \geq 0 & \end{cases}$$

Hence, it follows that X_o is an optimal solution to the following linear programming problem

$$\left.\begin{array}{ll} \max z = \nabla f(X_o)X & \\ \text{subject to} & \\ \nabla g_i(X_o)X \leq \nabla g_i(X_o)X_o & i \in I \\ X \geq 0 & \end{array}\right] \text{——— (P)}$$

Now problem (P) has a dual which by definition is

$$\left.\begin{array}{ll} \min Z = \sum_{i \in I} \lambda_i \left[\nabla g_i(X_o)X_o\right] & \\ \text{subject to} & \\ \sum_{i \in I} \lambda_i \left[\nabla g_i(X_o)\right]' \geq \left[\nabla f(X_o)\right]' & \\ \lambda_i \geq 0 & i \in I \end{array}\right] \text{——— (D)}$$

Now since problem (P) has an optimal solution, namely X_o, it follows, by the fundamental theorem of duality, that problem (D) has an optimal solution which we can denote by λ_i^o $i \in I$. If we take $\lambda_i^o = 0$ for $i \notin I$, the inequality constraints

$$\sum_{i\in I} \lambda_i^o [\nabla g_i(X_o)]' \geq [\nabla f(X_o)]' \qquad i \in I$$

can be written in the form

$$\nabla f(X_o) = \lambda_o' \nabla G(X_o) - \mu_o \tag{2.23}$$

where $\lambda_o = [\lambda_1^o, \ldots, \lambda_m^o] \geq 0$

$\mu_o = (\mu_1^o, \ldots, \mu_n^o) \geq 0$

which proves (i).

To prove (ii) we take the scalar product of (2.23) with X_o and obtain

$$\mu_o X_o = [\lambda_o' \nabla G(X_o)] X_o - \nabla f(X_o) X_o \ .$$

But by the fundamental theorem of duality max z = min Z, i.e.,

$$\nabla f(X_o) X_o = [\lambda_o' \nabla G(X_o)] X_o \ .$$

Hence $\mu_o X_o = 0$

Now

$g_i(X_o) = 0$ for $i \in I$

and $\lambda_i^o = 0$ for $i \notin I$

$\therefore$ $\lambda_o' G(X_o) = 0$

which proves (iii).

Q.E.D.

REMARK: THM. V gives a necessary condition for an optimal solution of a concave differentiable program which satisfies Hypothesis I or which is superconsistent. Later, in THM. IX, we shall prove that this condition is also sufficient.

THM. VI If for a given feasible solution X_o of Problem I there exists an m-component column vector $\lambda_o \geq 0$ and an n-component row vector $\mu_o \geq 0$, such that

(i) $\nabla f(X_o) = \lambda_o' \nabla G(X_o) - \mu_o$

(ii) $\mu_o X_o = 0$

(iii) $\lambda_o' G(X_o) = 0$,

then $[X_o,\lambda_o]$ solves Problem II.

PROOF: We are required to prove that $[X_o,\lambda_o]$ is a saddle point (global) of $F(X,\lambda)$ in the region $R^+ = \{[X,\lambda] \mid X \geq 0, \lambda \geq 0\}$.

Now, clearly $[X_o,\lambda_o] \in R^+ = \{[X,\lambda] \mid X \geq 0,\ \lambda \geq 0\}$ and so by THM. II a set of sufficient conditions, for $[X_o,\lambda_o]$ to be a saddle point (global) of $F(X,\lambda)$ in R^+, is:

$$\nabla_X F(X_o,\lambda_o) \leq 0\ ,\quad \nabla_X F(X_o,\lambda_o)X_o = 0 \tag{2.9}$$

$$\nabla_\lambda F(X_o,\lambda_o) \geq 0\ ,\quad \nabla_\lambda F(X_o,\lambda_o)\lambda_o = 0 \tag{2.10}$$

$$F(X,\lambda_o) \leq F(X_o,\lambda_o) + \nabla_X F(X_o,\lambda_o)(X-X_o) \quad \text{for all} \quad X \geq 0 \tag{2.11}$$

$$F(X_o,\lambda) \geq F(X_o,\lambda_o) + \nabla_\lambda F(X_o,\lambda_o)(\lambda-\lambda_o) \quad \text{for all} \quad \lambda \geq 0 \tag{2.12}$$

Now $F(X,\lambda) = f(X) - \lambda' G(X)$

$$\left.\begin{aligned} \therefore\ \nabla_X F(X_o,\lambda_o) &= \nabla f(X_o) - \lambda_o' \nabla G(X_o) \\ &= -\mu_o \qquad \text{by hypothesis (i)} \\ &\leq 0 \qquad \text{since} \quad \mu_o \geq 0 \end{aligned}\right| \quad \text{This verifies (2.9)}$$

But $\nabla_X(F(X_o,\lambda_o)X_o = -\mu_o X_o = 0$ by hypothesis (ii)

$$\left.\begin{aligned} &\text{Also } \nabla_\lambda F(X_o,\lambda_o) = -[G(X_o)]' \geq 0 \quad \text{since} \quad G(X_o) \leq 0 \\ &\text{But } \nabla_\lambda F(X_o,\lambda_o)\lambda_o = [G(X_o)]'\lambda_o = 0 \quad \text{(by hypothesis (iii))} \end{aligned}\right| \quad \text{This verifies (2.10)}$$

Now by THM.(I.1) we have that

$$G(X) \geq G(X_o) + \nabla G(X_o)(X-X_o) \quad \text{for any} \quad X \tag{2.24}$$

and by THM.(I.*1)

$$f(X) \leq f(X_o) + \nabla f(X_o)(X-X_o) \quad \text{for any} \quad x \tag{2.25}$$

Multiplying (2.24) by $-\lambda_o'$ and combining the resulting relation with (2.25) we obtain that for any $X \geq 0$

$$\begin{aligned} F(X,\lambda_o) &= f(X) - \lambda_o' G(X) \\ &\leq f(X_o) - \lambda_o' G(X_o) + (\nabla f(X_o) - \lambda_o' \nabla G(X_o))(X-X_o) \\ &= F(X_o,\lambda_o) + \nabla_X F(X_o,\lambda_o)(X-X_o)\ , \end{aligned}$$

which verifies (2.11).

Finally, relation (2.12) always holds because of linearity in λ, that is

$$\begin{aligned} F(X_o,\lambda) &= f(X_o) - \lambda' G(X_o) \\ &= f(X_o) - \lambda_o' G(X_o) + \lambda_o' G(X_o) - \lambda' G(X_o) \\ &= F(X_o,\lambda_o) + (\lambda-\lambda_o)' G(X_o) \\ &= F(X_o,\lambda_o) + [G(X_o)]'(\lambda-\lambda_o) \\ &= F(X_o,\lambda_o) + \nabla_\lambda F(X_o,\lambda_o)(\lambda-\lambda_o) , \end{aligned}$$

for any λ and in particular for any $\lambda \geq 0$.

Whence $[X_o,\lambda_o] \in R^+$ satisfies all the sufficient conditions for a saddle point (global) of $F(X,\lambda)$ in R^+, and hence $[X_o,\lambda_o]$ is a global saddle point of $F(X,\lambda)$ in R^+.

Q.E.D.

THM.VIII If for a given feasible solution X_o of Problem I there exists an m-component column vector $\lambda_o \geq 0$ such that $[X_o,\lambda_o]$ is a solution to Problem II, then X_o is an optimal solution to Problem I.

PROOF: Since $[X_o,\lambda_o]$ is a solution to Problem II, $[X_o,\lambda_o]$ is a global saddle point of $F(X,\lambda)$ in R^+. Therefore, by THM. II, we must have that

$$0 = \nabla_\lambda F(X_o,\lambda_o)\lambda_o = [-G(X_o)]'\lambda_o = -\lambda_o' G(X_o) \qquad (2.26)$$

Now, by definition of a saddle point (global) in R^+,

$$F(X,\lambda_o) \leq F(X_o,\lambda_o) \qquad \text{for all } X \geq 0 ,$$

that is

$$F(X,\lambda_o) = f(X) - \lambda_o' G(X) \leq f(X_o) - \lambda_o' G(X_o) = F(X_o,\lambda_o) \quad \forall\, X \geq 0 \qquad (2.27)$$

But $\lambda_o' G(X) \leq 0$ for all $X \in R_X = \{X | G(X) \leq 0, X \geq 0\}$

since $\lambda_o' \geq 0$. Furthermore by (2.26), $-\lambda_o' G(X_o) = 0$. Whence.

Hence by relation (2.27) we conclude that

$$f(X) \leq f(X_o) \qquad \text{for all } X \in R_X = \{X | G(X) \leq 0, X \geq 0\}$$

Q.E.D.

REMARK: THM. VII points out that the existence of a saddle point $[X_o, \lambda_o]$ of $F(X,\lambda)$ in the region R^+ implies the existence of an optimal solution to Problem I (and hence excludes the possibility of unboundedness in Problem I), provided the X component of the saddle point, i.e., X_o is a feasible solution to Problem I. Now, by THM. VI, if we have a feasible solution X to Problem I, for which there exists $\lambda \geq 0$ and $\mu \geq 0$ such that

$$\left.\begin{array}{ll} \text{(i)} & \nabla f(X) = \lambda' \nabla G(X) - \mu \\ \text{(ii)} & \mu X = 0 \\ \text{(iii)} & \lambda' G(X) = 0 \end{array}\right] \qquad (2.28)$$

then $[X,\lambda]$ is a saddle point (global) to $F(X,\lambda)$ in R^+; whence if we can find a feasible solution X_* to Problem I for which there are vectors $\lambda_* \geq 0$, $\mu_* \geq 0$ such that the triple X_*, λ_*, μ_* satisfies (2.28) then we immediately know that Problem I has an optimal solution (and so cannot be unbounded), namely X_*. This provides a sufficient criterion for deciding whether or not the differentiable program given in Problem I has an optimal solution and a criterion for finding an optimal solution to Problem I. Note also that THM. VI, THM. VII, and all the conclusions of this remark have been deduced without assuming either that Problem I is superconsistent or that it satisfies Hypothesis I.

<u>THM.VIII</u> If the concave differentiable program given in Problem I satisfies Hypothesis I, then a necessary and sufficient condition for a feasible solution X_o of Problem I to be optimal is that there exist an m-component column vector $\lambda_o \geq 0$ such that $[X_o, \lambda_o]$ solves Problem II.

PROOF: Necessity. Let X_o be an optimal solution to Problem I, then by THM. V there exists an m-component column vector $\lambda_o \geq 0$ and an n-component row vector $\mu_o \geq 0$ such that

$$\text{(i)}\quad \nabla f(X_o) = \lambda_o' \nabla G(X_o) - \mu_o$$

$$\text{(ii)}\quad \mu_o X_o = 0$$

$$\text{(iii)}\quad \lambda_o' G(X_o) = 0$$

Hence by THM. VI, $[X_o, \lambda_o]$ solves Problem II.

Sufficiency. Follows immediately from THM. VII. Q.E.D.

THM. IX If the concave differentiable program given in Problem I satisfies Hypothesis I, then a necessary and sufficient condition for a feasible solution X_o of Problem I to be optimal is that there exists an m-component column vector $\lambda_o \geq 0$ and an n-component row vector $\mu_o \geq 0$ such that:

$$\text{(i)}\quad \nabla f(X_o) = \lambda_o' \nabla G(X_o) - \mu_o$$

$$\text{(ii)}\quad \mu_o X_o = 0$$

$$\text{(iii)}\quad \lambda_o' G(X_o) = 0$$

PROOF: Necessity. Proved in THM. V.

Sufficiency. Suppose $X_o \in R_X = \{X \mid G(X) \leq 0, X \geq 0\}$ and that there exists $\lambda_o \geq 0$ and $\mu_o \geq 0$ such that (i), (ii) and (iii) are satisfied. Then by THM. VI $[X_o, \lambda_o]$ solves Problem II. Hence by THM. VII X_o is an optimal solution to Problem I.

Q.E.D.

(2.4) Concave Differentiable Programs with Linear Constraints

If α is an n-component row vector belonging to E_n then the linear function $g(X) = \alpha X$ is a convex function in E_n (see THM. I Chapter III). Hence if in Problem I the constraints $g_i(X) \leq 0$ $i = 1, \ldots, m$ are linear we will continue to have a concave differentiable program. This particular type of concave differentiable program plays a most important role in the theory of quadratic programming and so we introduce the following definition.

DEFN. <u>Concave Differentiable Program with m Linear Constraints.</u>

A concave differentiable program with m linear constraints is a concave differentiable program with exactly m convex constraints $g_i(X) \leq 0$, $i = 1,\ldots,m$, which are assumed to be linear.

<u>THM. X</u> Suppose that in the concave differentiable program given in Problem I, the constraints $g_i(X) \leq 0$, $i = 1,\ldots,m$, are linear and suppose that X_o is an optimal solution to Problem I. Then,

$$\nabla f(X_o)(X-X_o) \leq 0$$

for all X satisfying the inequalities

$$\nabla g_i(X_o)(X-X_o) \leq 0 \qquad i \in I = \{i|g_i(X_o) = 0\}$$
$$X \geq 0$$

PROOF: Since the constraints $g_i(X) \leq 0$, $i = 1,\ldots,m$, are linear we can write

$$g_i(X) = \alpha_i X \qquad i = 1,\ldots,m$$

where α_i is an n-component row vector of constants. Hence $\nabla g_i(X) = \alpha_i$, $i = 1,\ldots,m$. Now suppose there exists an arbitrary $X \in R_o = \{X|X \geq 0,$ $\nabla g_i(X_o)(X-X_o) \leq 0 \ \ i \in I\}$ such that $\nabla f(X_o)(X-X_o) > 0$. Then

$$0 \geq \nabla g_i(X_o)(X-X_o) = \alpha_i(X-X_o)$$
$$= \alpha_i X \qquad \text{for all } i \in I .$$

Hence X is a feasible point of the feasible domain $R_1 = \{X|X \geq 0,$ $g_i(X) = \alpha_i X \leq 0 \ \ i \in I\}$. Also $X_o \in R_1$. Thus, since R_1 is convex, for sufficiently small $\lambda > 0$ the point $X_o + \lambda(X-X_o)$ still lies in R_1.

Therefore, since $\nabla f(X_o)(X-X_o) > 0$, it follows by THM.(V^*.I), that

$$f(X_o+\lambda(X_1-X_o)) > f(X_o),$$

for sufficiently small λ.

Hence X_o is not an optimal solution to the problem

$$\left.\begin{array}{ll} \max f(X) & \\ \text{subject to} & \\ \alpha_i X \leq 0 & i \in I \\ X \geq 0 & \end{array}\right| \qquad (2.29)$$

But this is a contradiction since X_o is an optimal solution to Problem I and hence by THM. I, it must be a solution to (2.29). Whence for all arbitrary $X \in R_o$ we must have that $\nabla f(X_o)(X-X_o) < 0$

Q.E.D.

REMARK: THM. X shows us that a concave differentiable program with linear constraints automatically satisfies Hypothesis I and hence we have at once the following theorem:

<u>THM. XI</u> Given a concave differentiable program with m linear constraints, then a necessary and sufficient condition for a feasible solution X_o of this program to be optimal is that there exists an m-component column vector $\lambda_o \geq 0$ and an n-component row vector $\mu_o \geq 0$ such that:

$$\text{(i)}\ \nabla f(X_o) = \lambda_o' \nabla G(X_o) - \mu_o$$

$$\text{(ii)}\ \mu_o X_o = 0$$

$$\text{(iii)}\ \lambda_o' G(X_o) = 0\ .$$

PROOF: A concave differentiable program with m linear constraints has the form of Problem I and satisfies Hypothesis I (THM. X). Hence result follows immediately from THM. IX.

Q.E.D.

REMARK: (i) In THM. XI note that the number of components in the column vector $\lambda_o = [\lambda_1^o,\ldots,\lambda_m^o] \geq 0$ is equal to the number of linear constraints in the concave differentiable program. So if in the hypothesis of THM. XI we have a concave differentiable program with 2m constraints instead of one with m constraints, then the corresponding λ_o will have 2m components.

(ii) The components $\lambda_1^o,\ldots,\lambda_m^o$ of λ_o are called Lagrange Multipliers.

(2.5) Characterisation of an Optimal Solution of a Convex Differentiable Program.

In section (2.3) we derived the theory concerning optimal solutions of concave differentiable programs. Whenever we are met with a convex differentiable program, namely,

$$\left.\begin{array}{l} \text{minimize} \quad f(X) \\ \text{subject to} \\ \qquad g_i(X) \leq 0 \qquad i = 1,\ldots,m \\ \qquad X \geq 0 \\ \text{given that:} \\ \quad \text{(i) } f(X) \in C' \text{ and } f(X) \text{ is a convex function on } E_n \\ \quad \text{(ii) } g_i(X) \in C' \text{ and } g_i(X) \text{ is a convex function on } E_n \text{ for } i=1,\ldots,m, \end{array}\right] \quad (2.30)$$

we transform it, with the help of the relation

$$\text{minimum} \quad f(X) = -\text{maximum} \quad -f(X) ,$$

into the equivalent concave differentiable program

$$\left.\begin{array}{l} \text{maximize} \quad -f(X) \\ \text{subject to} \quad g_i(X) \leq 0, \quad i = 1,\ldots,m, \quad X \geq 0 \end{array}\right] \quad (2.31)$$

and then apply all the theory derived in sections (2.3) and (2.4) to problem (2.31).

Chapter III

"Quadratic Programs"

(3.1) Quadratic Programs in Normal Form.

DEFN. Quadratic Program in Normal Form - is a problem of the type

$$\left.\begin{array}{l} \text{minimize or maximize } f(X) = CX + X'DX \\ \text{subject to} \\ \qquad AX = b \\ \qquad X \geq 0 \end{array}\right] \qquad (N^2)$$

where:

(i) $C = (c_1,\ldots,c_n)$ is an n-component row vector

(ii) $X = [x_1,\ldots,x_n]$ is an n-component column vector (of variables)

(iii) $b = [b_1,\ldots,b_m]$ ($\neq 0$) is an m-component column vector.

(iv) $m \leq n$

(v) $D = \|d_{ij}\|$ is an $n\times n$ symmetric matrix

(vi) $A = \|a_{ij}\|$ is an $m\times n$ matrix of rank m.

DEFN. Negative Definite Quadratic Form. A quadratic form $X'DX$ is said to be negative definite if

$$X'DX < 0 \quad \text{for all} \quad X \neq 0 .$$

DEFN. Negative Semi-Definite Quadratic Form. A quadratic form $X'DX$ is said to be negative semi-definite if

$$X'DX \leq 0 \quad \text{for all} \quad X.$$

REMARK: A quadratic form $X'DX$ is said to be positive definite if $-X'DX$ is negative definite, and is said to be positive semi-definite if $-XDX$ is negative semi-definite.

DEFN. Indefinite Quadratic Form. A quadratic form $X'DX$ is said to be indefinite if it is negative for some points X and positive for others.

DEFN. Quadratic Program in Normal Form I is a quadratic program in Normal Form where:

(i) the objective function $f(X) = CX + X'DX$ is to be maximized,

(ii) the quadratic form $X'DX$ is negative definite.

DEFN. Quadratic Program in Normal Form II is a quadratic program in Normal Form where:

(i) the objective function $f(X) = CX + X'DX$ is to be maximized,

(ii) the quadratic form $X'DX$ is negative semi-definite.

DEFN. Quadratic Program in General Normal Form is a quadratic program in Normal Form where:

(i) the objective function $f(X) = CX + X'DX$ is to be maximized,

(ii) the quadratic form $X'DX$ is indefinite.

REMARK: Note that both in Normal Form I and Normal Form II the elements of A, b and C are arbitrary, but those of D are not. At present there is no suitable method for solving a quadratic program in Normal Form with an arbitrary (indefinite) symmetric matrix D. Consequently, we will restrict ourselves to the study of algorithms for the solution of quadratic programs either in Normal Form I or in Normal Form II. (See Chapter IV - Wolfe's Method and Chapter VI - Dantzig's Method).

(ii) A function $f(X)$ of the type $f(X) = f_1(X) + f_2(X)$ where $f_1(X)$ is a linear form and $f_2(X)$ is a quadratic form is called <u>quadratic function</u>.

(3.2) <u>Quadratic Programs in Slack Form.</u>

DEFN. <u>Quadratic Program in Slack Form</u> - is a problem of the type

minimise or maximise $f(X) = CX + X'DX$

subject to

$$AX \leq b$$

$$X \geq 0 ,$$

where C, X, b, m, n, D and A are as defined in section (3.1).

DEFN. <u>Quadratic Program in Slack Form I</u> - is a quadratic program in Slack Form where:

(i) the objective function $f(X) = CX + X'DX$ to be maximised

(ii) the quadratic form $X'DX$ is negative definite.

DEFN. <u>Quadratic Program in Slack Form II</u> - is a quadratic program in Slack Form where:

(i) the objective function $f(X) = CX + X'DX$ is to be maximised

(ii) the quadratic form $X'DX$ is negative definite

DEFN. <u>Quadratic Program in General Slack Form</u> - is a quadratic program in Slack Form where:

(i) the objective function $f(X) = CX+X'DX$ is to be maximised

(ii) the quadratic form $X'DX$ is indefinite.

REMARK: (i) Note that again, both in Slack Form I and Slack Form II the elements of A, b′ and C are arbitrary, but those of D are not. At present there is no suitable method for solving a quadratic program in Slack Form with an arbitrary (indefinite) symmetric matrix D. Consequently, we will restrict ourselves to the study of algorithms for the solution of quadratic programs either in Slack Form I or in Slack Form II (see Chapter V - Frank and Wolfe Method).

(3.3) Other Forms of Quadratic Programs.

DEFN. Quadratic Program in Mixed Form - is a problem in which we want to minimise or maximise a quadratic function $f(X) = CX + X'DX$, subject to the "mixed" constraints

$$\begin{array}{ll} \alpha_i X = b_i & i = 1,\ldots,k \\ \alpha_i X \leq b_i & i = k+1,\ldots,\ell \\ \alpha_i X \geq b_i & i = \ell+1,\ldots,m \\ x_j \geq 0 & j = 1,\ldots,p \\ x_j \leq 0 & j = p+1,\ldots,q \\ x_j \text{ unrestricted} & j = q+1,\ldots,n \end{array}$$

where:

(i) α_i, $i = 1,\ldots,m$ denotes an n-component row vector of scalars.

(ii) The rank of the matrix $A = \begin{bmatrix} \alpha_1 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_m \end{bmatrix}$ is m.

(iii) C, X and D are as defined in section (3.1).

(iv) b_i, $i = 1,\ldots,m$, is a scalar.

DEFN. Quadratic Program in Mixed Form I - is a quadratic program in Mixed Form where:

(i) the objective function $f(X) = CX + X'DX$ is to be maximised

(ii) the quadratic form $X'DX$ is negative definite.

DEFN. Quadratic Program in Mixed Form II - is a quadratic program in Mixed Form where:

(i) the objective function $f(X) = CX + X'DX$ is to be maximised

(ii) the quadratic form $X'DX$ is negative semi-definite.

DEFN. Quadratic Program in General Mixed Form - is a quadratic program in Mixed Form where:

(i) the objective function $f(X) = CX + X'DX$ is to be maximised

(ii) the quadratic form $X'DX$ is indefinite.

A quadratic program in Mixed Form I (II) can be reduced to a quadratic program in Normal Form I (II) or in Slack Form I(II) by the use of operations (1), (2) and (5) (below) in the first case or by the use of operations (1), (2) (3) and (4) in the second case. So, any problem of maximizing a quadratic objective function subject to linear constraints can be reduced to one of the normal forms ($N^2.I$ or $N^2.II$) or to one of the slack forms ($S^2.I$ or $S^2.II$) by the use of operations (1), (2), (3), (4) and (5) (where necessary), provided the objective function is concave or strictly concave. Similarly, any problem of minimizing a quadratic objective function $f(X)$ subject to linear constraints can be reduced to one of the normal forms ($N^2.I$ or $N^2.II$) or to one of the slack forms ($S^2.I$ or $S^2.II$) by the use of operations (1), (2), (3), (4), (5) and (6), but in this case the objective function must be convex or strictly convex (this is because of operation (6) and because $-f(X)$ is concave if and only if $f(X)$ is convex).

Operation (1): A variable of negative sign, x, may always be replaced by a non-negative variable x^+. It suffices to make the change of variable

$$x^+ = -x .$$

Operation (2): A variable of arbitrary sign, x, may always be replaced by two non-negative variables x^+ and x^-. It suffices to make the change of variables

$$x = x^+ - x^-$$

where

$$x^+ = \text{maximum } [0, x]$$

$$x^- = \text{maximum } [0, -x]$$

Operation (3): Every equation $\alpha_i X = b_i$ may be replaced by the two inequalities

$$\alpha_i X \leq b_i$$

$$-\alpha_i X \leq -b_i$$

Operation (4): Every inequality $\alpha_i X \geq b_i$ may be replaced by the inequality $-\alpha_i X \leq -b_i$

Operation (5): Every inequality $\alpha_i X \geq b_i$ or $\alpha_i X \leq b_i$, may be replaced, respectively, by the relations

$$\text{(i)} \text{———} \begin{cases} \alpha_i X - y_i = b_i \\ \qquad y_i \geq 0 \end{cases}$$

or

$$\text{(ii)} \text{———} \begin{cases} \alpha_i X + y_i = b_i \\ \qquad y_i \geq 0 \end{cases} ,$$

obtained by subtraction or addition of a supplementary non-negative variable, called the surplus variable in case (i) or slack variable in case (ii), which is given a zero coefficient in the quadratic form to be optimised. Note that, since in all the solution methods to be presented in the sequel, the non-negativity constraints may be taken into account without increasing the volume of calculations (this will be easily seen once the descriptions of these methods have been given), the preceding transformation may be interpreted as replacing an inequality by an equality, at the cost of adding one variable.

Operation (6): Using the relation

$$\text{minimum } f(X) = -\text{ maximum } [-f(X)],$$

in which $f(X)$ represents the quadratic function to be optimised, any quadratic programming problem can always be expressed as a maximization (or a minimization) problem

Note that the two formulations of quadratic programming problems, Normal Form and Slack Form, only provide a mathematical simplification. In fact they are equivalent since the "modified" constraints of Problem ($N^2_.$) can be reduced to the form of Problem ($S^2_.$), by using Operation (3) above, and the "modified" constraints of Problem ($S^2_.$) can be reduced to the form of Problem ($N^2_.$), by using Operation (5) above. More specifically, if one wants to express Problem ($N^2_.$) in the form of Problem ($S^2_.$) one sets

$$A^* = \begin{bmatrix} A \\ -A \end{bmatrix}, \qquad b^* = \begin{bmatrix} b \\ -b \end{bmatrix},$$

and then Problem ($N^2_.$) is equivalent to the problem.

$$\text{minimize of maximize } \quad f(X) = CX + X'DX$$

subject to

$$A^*X \leq b^*$$

$$X \geq 0 .$$

Conversely, if one wants to express Problem $(S_{\bullet}^{2})$ in the form of Problem $(N_{\bullet}^{2})$, one introduces a slack vector $Y \geq 0$ and replaces the system of constraints $AX \leq b$ by $AX + Y = b$, $Y \geq 0$. Problem $(S_{\bullet}^{2})$ is then equivalent to the problem:

$$\text{minimize or maximize} \quad f(X^*) = C^*X^* + (X^*)'D^*X^*$$

subject to

$$A^{**}X^* = b$$

$$X^* \geq 0$$

where

$$X^* = \begin{bmatrix} X \\ Y \end{bmatrix}, \quad A^{**} = (A, I_m), \quad D^* = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}, \quad C^* = \begin{bmatrix} C \\ 0 \end{bmatrix}$$

In the following, we will consider whichever form best suits the particular method under discussion.

Note also that the classification of quadratic programs (in Normal, Slack or Mixed Forms) into a Form I, a Form II and a General Form, is motivated by the fact that at present there is no suitable method for solving a general quadratic programming problem with an arbitrary (indefinite) quadratic form $X'DX$. The solution methods available only work if $X'DX$ is semi-definite (positive or negative) or strictly definite (positive or negative). As a matter of fact quite a few of the available methods need the strict definiteness of $X'DX$ in order to obtain an optimal solution to the quadratic Program (see App. III.1).

Finally we end this section by pointing out that:

(i) In quadratic programs the qualifications, Normal, Slack and Mixed refer to the type of constraints:

Normal → {=}

Slack → {$\leq$}

Mixed → {=, $\leq$, $\geq$} .

(ii) In quadratic programs the qualifications, Form I, Form II and General Form refer to the optimization of the objective function $f(X) = CX + X'DX$ (i.e. maximization instead of minimization) and to the quadratic form $X'DX$:

Form I → maximization of $f(X) = CX + X'DX$, with $X'DX$ negative definite

Form II → " " " " , with $X'DX$ negative semi-definite

General Form → " " " " , with $X'DX$ indefinite

(iii) From now on whenever we use the terminology "quadratic programs" it should be understood that we are either referring to quadratic programs in Normal Form (I, II or General) or to quadratic programs in Slack Form (I, II or General) or both.

(iv) Results will only be derived for quadratic programs in Normal Form (I or II) or in Slack Form (I or II) and the reader may apply any results derived for these programs to any other quadratic programming problem (as long as it is not in General Form) by making use of the above transformations.

(3.4) Properties of the Quadratic Function $f(X) = CX + X'DX$.

THM. I The linear form CX is a concave function in E_n.

PROOF: (See following page)

PROOF: Let X_1 and X_2 be any two points in E_n such that $X_1 \neq X_2$. Let $v(X) = CX$ then

$$v(\lambda X_2 + (1-\lambda)X_1) = C(\lambda X_2 + (1-\lambda)X_1)$$
$$= \lambda CX_2 + (1-\lambda)CX_1 \ ,$$

for any λ, and in particular for any $\lambda \in [0,1]$.

Q.E.D.

REMARK: (i) From the proof of the above theorem it immediately follows that the linear form $v(X) = CX$ is also a convex function in E_n.

(ii) Note also that $CX + d$ where d is any real number is a convex function in E_n.

THM. II A negative semi-definite quadratic form $z(X) = X'DX$ in E_n is a concave function in E_n.

PROOF: Let X_1 and X_2 be any two points in E_n such that $X_1 \neq X_2$. Let $z(X) = X'DX$ and consider

$$\begin{aligned} z(\lambda X_2+(1-\lambda)X_1) &= [\lambda X_2 + (1-\lambda)X_1]'D[\lambda X_2 + (1-\lambda)X_1] \\ &= \lambda^2 X_2'DX_2+\lambda(1-\lambda)X_2'DX_1+\lambda(1-\lambda)X_1'DX_2+(1-\lambda)^2X_1'DX_1 \\ &= \lambda^2 X_2'DX_2+\lambda(1-\lambda)(X_1'DX_1+X_2'DX_2-[X_2-X_1]'D[X_2-X_1])+(1-\lambda)^2X_1'DX_1 \\ &\geq \lambda^2 X_2'DX_2+\lambda(1-\lambda)(X_1'DX_1+X_2'DX_2)+(1-\lambda)^2X_1'DX_1 \ , \qquad (3.0) \end{aligned}$$

for $0 \leq \lambda \leq 1$, since then $-\lambda(1-\lambda)[X_2-X_1]'D[X_2-X_1] \geq 0$.

Therefore

$$\begin{aligned} z(\lambda X_2 + (1-\lambda)X_1) &\geq \lambda X_2'DX_2 + (1-\lambda)(\lambda+1-\lambda)X_1'DX_1 \quad \text{for } 0 \leq \lambda \leq 1 \\ &= \lambda X_2'DX_2 + (1-\lambda)X_1'DX_1 \quad \text{for } 0 \leq \lambda \leq 1 \\ &= \lambda z(X_2) + (1-\lambda)z(X_1) \quad \text{for } 0 \leq \lambda \leq 1 \end{aligned}$$

Q.E.D.

THM. III A negative definite quadratic form $X'DX$ in E_n is a strictly concave function in E_n.

PROOF: Follows precisely the same lines as THM. II the only difference being that we can write relation (3.0) as a strict inequality since now we have that $z(X)$ is negative definite and hence

$$-\lambda(1-\lambda)[X_2-X_1]'D[X_2-X_1] > 0 \quad \text{for} \quad 0 \leq \lambda \leq 1$$

THM. IV If $z(X) = X'DX$ is a negative semi-definite quadratic form in E_n then the function $f(X) = CX + X'DX$ is a concave function in E_n.

PROOF: By THM. I CX is a concave function in E_n. By THM. II $z(X) = X'DX$ is a concave function in E_n. Hence result follows at once from THM.(II*.1)

Q.E.D.

THM. V If $X'DX$ is a negative definite quadratic form in E_n then the function $f(X) = CX + X'DX$ is a strictly concave function in E_n.

PROOF: Let X_1 and X_2 be any two points in E_n such that $X_1 \neq X_2$. Let $v(X) = CX$ and $z(X) = X'DX$ then $f(X) = v(X) + z(X)$, and so

$$\begin{aligned} f(\lambda X_2 + (1-\lambda)X_1) &= v(\lambda X_2+(1-\lambda)X_1)+z(\lambda X_2+(1-\lambda)X_1) \\ &\geq \lambda v(X_2)+(1-\lambda)v(X_1)+z(\lambda X_2+(1-\lambda)X_1) \quad \forall\ , 0\leq\lambda\leq 1 (\text{by THM.I}) \\ &> \lambda v(X_2)+(1-\lambda)v(X_1)+\lambda z(X_2)+(1-\lambda)z(X_1) \forall\lambda, 0\leq\lambda\leq 1 (\text{by THM.III}) \end{aligned}$$

Hence

$$\begin{aligned} f(\lambda X_2 + (1-\lambda)X_1) &> \lambda(v(X_2)+z(X_2))+(1-\lambda)(v(X_1)+z(X_1)) \quad \forall\lambda, 0\leq\lambda\leq 1 \\ &= \lambda f(X_2) + (1-\lambda)f(X_1) \quad \forall\lambda, 0\leq\lambda\leq 1 \end{aligned}$$

Q.E.D.

THM. VI. If the quadratic function $f(X) = CX + X'DX$ (not necessarily concave or convex) is bounded above on the non-empty set $\mathbb{K}_S = \{X | AX \leq b, X \geq 0\}$, then it attains its supremum (maximum) on $\mathbb{K}_S$.

PROOF: (See Appendix III.2)

THM. VII. If the quadratic function $f(X) = CX + X'DX$ (not necessarily concave or convex) is bounded above on the non-empty convex set $\mathbb{K}_N = \{X | AX = b, X \geq 0\}$, then it attains its supremum (maximum) on $\mathbb{K}_N$.

PROOF: (See Appendix III.2).

REMARKS: (i) THM. VI and VII show us that if the objective function of a quadratic program (any type) is bounded above on the feasible domain then the program has at least one optimal feasible solution. Hence with quadratic programs the boundedness of the objective function over the feasible domain is sufficient to imply solvability. Note however that this property does not apply to concave differentiable programs in general:- the strictly concave function $-e^{-x}$ which is bounded above by 0 in the domain $x \geq 0$ but does not assume its supremum there, serves as an illustrative example.

(ii) Note that THM. VI and VII remain true if the function $f(X)$ is simply a linear form, i.e., if $f(X) = CX$, for then we can regard $f(X)$ as the quadratic function $f(X) = CX + X'DX$, where $D = 0$.

<u>THM. VIII</u> Let $f(X) = CX + X'DX$. If the quadratic form $X'DX$ is negative definite in E_n then

$$f(X) \to -\infty \quad \text{as} \quad |X| \to \infty$$

PROOF: Consider any point X lying on the hypersphere $H_t = \{X \mid |X| = t>0\}$, of radius t with centre at the origin. Then $X = tR$, where R is a point on the hypersphere $H_1 = \{X \mid |X| = 1\}$, of radius unity with center at the origin. And so $X'DX = t^2R'DR$. Hence $R'DR$ is negative definite in E_n since $X'DX$ is negative definite in E_n.

Now let the maximum of $R'DR$ over the unit hypersphere H_1 be taken on at R_o. Then $R_o'DR_o = d$, say, where $d < 0$. Therefore

$$X'DX = t^2R'DR \leq t^2d \qquad \forall\, X \in H_t.$$

Thus as $|X| = t \to \infty$, $X'DX \to -\infty$ since $d < 0$, i.e.,

$$\lim_{|X| \to -\infty} X'DX = -\infty. \tag{3.1}$$

Now if $X \neq 0$ we can write $f(X)$ in the form

$$f(X) = X'DX \left[1 + \frac{CX}{X'DX}\right]$$

Let the maximum of $\frac{CR}{R'DR}$ over the unit hypersphere H_1 be taken on at R_1 then,

$$\left[\frac{CX}{X'DX}\right] = \left[\frac{tCR}{t^2R'DR}\right]$$

$$= \frac{1}{t}\left[\frac{CR}{R'DR}\right]$$

$$\leq \frac{1}{t}\left[\frac{CR_1}{R_1'DR_1}\right] \qquad \forall\, X \in H_t\,.$$

(Note that the value of $\frac{CR_1}{R_1'DR_1}$ is finite since $R_1'DR_1 \neq 0$).

Therefore as $|X| = t \to \infty$, $\frac{CX}{X'DX} \to 0$ and so

$$f(X) = X'DX\left[1 + \frac{CX}{X'DX}\right] \to -\infty \quad \text{as} \quad |X| = t \to \infty$$

since from (3.1) we have that $\lim_{|X|\to\infty} X'DX = -\infty$.

Whence

$$f(X) \to -\infty \quad \text{as} \quad |X| \to \infty$$

Q.E.D.

(3.5) Quadratic Programs in Normal Form as Concave Differentiable Programs With Linear Constraints.

In problem $(N_\circ^2)$ if we let the rows of the matrix $A = \|a_{ij}\|$ be denoted by $\alpha_i = (a_{i1}, a_{i2}, \ldots, a_{in})$, $i = 1,\ldots,m$, and the columns of A be denoted by $a_j = [a_{1j}, a_{2j}, \ldots, a_{mj}]$ $j = 1,\ldots,n$, then the set of constraints $AX = b$ can be written either in the form

$$\alpha_i X - b_i = 0 \qquad i = 1,\ldots,m \tag{3.2}$$

or in the form

$$\sum_{j=1}^{n} a_j x_j - b = 0 \tag{3.3}$$

Furthermore

$$\{X|\alpha_i X - b_i = 0, i = 1,\ldots,m\} = \{X|\alpha_i X - b_i \leq 0,\ \alpha_i X - b_i \geq 0,\ i = 1,\ldots,m\}$$

so that the set of equality constraints (3.2) can be written as a set of inequality constraints, namely

$$\alpha_i\ X - b_i \leq 0 \qquad i = 1,\ldots,m$$

$$\alpha_{i-m} X - b_{i-m} \leq 0 \qquad i = m+1,\ldots,2m\ ,$$

where $\alpha_{i-m} = (a_{i-m,1},\ldots,a_{i-m,n})$ $i = m+1,\ldots,2m$.

Therefore, we can write a quadratic program in Normal Form I as

$$\left.\begin{array}{ll}
\text{maximize} & f(X) = CX + X'DX \\
\text{subject to} & \\
& g_i(X) = \alpha_i X - b_i \leq 0 \quad i = 1,\ldots,m \quad (AX-b \leq 0) \\
& g_i(X) = -(\alpha_{i-m}X - b_{i-m}) \leq 0 \quad i = m+1,\ldots,2m \quad (-(AX-b) \leq 0) \\
& X \geq 0 \\
\text{given that the quadratic form } X'DX \text{ is negative definite} &
\end{array}\right] (N^2I)$$

Similarly we can write a quadratic program in Normal Form II as

$$\left.\begin{array}{ll}
\text{maximize} & f(X) = CX + X'DX \\
\text{subject to} & \\
& g_i(X) = \alpha_i X - b_i \leq 0 \quad i = 1,\ldots,m \quad (AX-b \leq 0) \\
& g_i(X) = -(\alpha_{i-m}X - b_{i-m}) \leq 0 \quad i = m+1,\ldots,2m \quad (-(AX-b) \leq 0) \\
& X \geq 0\,, \\
\text{given that the quadratic form } X'DX \text{ is \underline{negative semi-definite}.} &
\end{array}\right] (N^2II)$$

Now by THM. IV and V the objective function $f(X)$ of problems (N^2I) and (N^2II) is concave in E_n. Furthermore $f(X)$ clearly $\in C'$, in both cases. Hence problems (N^2I) and (N^2II) are in reality concave differentiable programs with <u>2m</u> linear constraints (compare with definition given in section (2.4) of Chapter II). Therefore by THM.(XI.2) we have that a necessary and sufficient condition for a feasible solution X to problems (N^2I) and (N^2II) to be optimal, is that there exists a <u>2m-component</u> column vector $\phi \geq 0$ (here we prefer to use ϕ instead of our usual λ) and an n-component row vector $\mu \geq 0$ such that

(i) $\nabla f(X) = \phi'\nabla G(X) - \mu$

(ii) $\mu X = 0$

(iii) $\phi' G(X) = 0$.

Now $$\frac{\partial}{\partial x_j} f(X) = c_j + 2 \sum_{i=1}^{n} x_i d_{ij} \qquad j = 1,\dots,n$$

$$\frac{\partial}{\partial x_j} g_i(X) = a_{ij} \qquad i = 1,\dots,m \text{ and } j = 1,\dots,n$$

$$\frac{\partial}{\partial x_j} g_i(X) = - a_{i-m,j} \qquad i = m+1,\dots,2m \text{ and } j = 1,\dots,n$$

Therefore

$$\nabla f(X) = C + 2X'D \tag{3.4}$$

$$\nabla g_i(X) = \alpha_i \qquad i = 1,\dots,m$$

$$\nabla g_i(X) = -\alpha_{i-m} \qquad i = m+1,\dots,2m$$

Thus

$$\begin{aligned} G(X) &= [\nabla g_1(X),\dots,\nabla g_m(X); \nabla g_{m+1}(X),\dots,\nabla g_{2m}(X)] \\ &= [\alpha_1,\dots\dots, \alpha_m ; -\alpha_1 ,\dots,-\alpha_m] \\ &= \begin{bmatrix} A \\ -A \end{bmatrix} \end{aligned} \tag{3.5}$$

Also $$G(X) = \begin{bmatrix} AX - b \\ -(AX - b) \end{bmatrix} \tag{3.6}$$

If we now partition the 2m-component column vector $\phi \geq 0$ into two m-component column vectors $\omega \geq 0$ and $\eta \geq 0$, so that $\phi = [\omega,\eta]$, we obtain from (3.4), (3.5) and (3.6), that a necessary and sufficient condition for a feasible solution X of problems $(Q^2_{\cdot}I)$ and $(Q^2_{\cdot}II)$ to be optimal, is that there exists two m component column vectors $\omega \geq 0$, $\eta \geq 0$ and an n component row vector $\mu \geq 0$ such that

$$\begin{aligned} &\text{(i)} \quad C + 2X'D = \omega'A - \eta'A - \mu \\ &\text{(ii)} \quad \mu X = 0 \\ &\text{(iii)} \quad \omega'(AX-b) - \eta'(AX-b) = 0 \end{aligned} \tag{3.7}$$

If we next let $\mu = V'$ then V will be an n-component column vector and equations (i) and (ii) of (3.7) become respectively

(i) $2DX - A'\omega + A'\eta + V = -C'$ because $D' = D$

(ii) $V'X = 0$.

Since $AX - b = 0$ equation (iii) of (3.7) becomes simply

(iii) $AX = b$.

Hence we conclude that a necessary and sufficient condition for a feasible solution X of problems $(Q^2_{\cdot}I)$ and $(Q^2_{\cdot}II)$ to be optimal is that there exists two m component column vectors $\omega \geq 0$ and $\eta \geq 0$ and an n component column vector $v \geq 0$ such that

$$\left.\begin{array}{lr} \text{(i) } 2DX - A'\omega + A'\eta + V = -C' & (3.8.a) \\ \text{(ii) } V'X = 0 & (3.8.b) \\ \text{(iii) } AX = b & (3.8.c) \end{array}\right] \text{———} (3.8)$$

Finally note that equation (3.8.a) can be written as

$$2DX - A'(\omega-\eta) + V = -C'$$

Thus if we put $\lambda = \omega-\eta$, then λ will be an m-component column vector unrestricted in sign since $\omega \geq 0$ and $\eta \geq 0$. So, alternatively, we can state that a necessary and sufficient condition for a feasible solution X of problems $(Q^2_{\cdot}I)$ and $(Q^2_{\cdot}II)$ to be optimal is that there exists an m-component column vector λ, <u>unrestricted in sign</u>, and an n component column vector $V \geq 0$ such that

$$\left.\begin{array}{lr} \text{(i) } 2DX - A'\lambda + V = -C' & (3.9.a) \\ \text{(ii) } V'X = 0 & (3.9.b) \\ \text{(iii) } AX = b \text{ .} & (3.9.c) \end{array}\right] \text{———} (3.9)$$

(3.6) <u>Quadratic Programs in Slack Form as Concave Differentiable Programs with Linear Constraints.</u>

Adopting the notation of section (3.5) we can write a quadratic program in Slack Form I as

$$\text{maximize} \quad f(X) = CX + X'DX$$

subject to

$$g_i(X) = \alpha_i(X) - b_i \leq 0 \qquad i = 1,\ldots,m \quad (AX-b \leq 0)$$

$$X \geq 0$$

given that the quadratic form $X'DX$ is <u>negative definite</u>. $(S^2.I)$

Similarly we can write a quadratic program in Slack Form II as

$$\text{maximize} \quad f(X) = CX + X'DX$$

subject to

$$g_i(X) = \alpha_i(X) - b_i \leq 0 \qquad i = 1,\ldots,m \quad (AX-b \leq 0)$$

$$X \geq 0$$

given that the quadratic form $X'DX$ is negative semi-definite $(S^2.I)$

Now by THM. IV and V the objective function $f(X)$ of Problems $(S^2.I)$ and $(S^2.II)$ is concave in E_n. Furthermore $f(X)$ clearly $\in C'$, in both cases. Hence problems $(S^2.I)$ and $(S^2.II)$ are in reality concave differentiable programs with m linear constraints (compare with definition given in section (2.4) of Chapter II). Therefore by THM. (XI.2) we have that a necessary and sufficient condition for a feasible solution X to problems $(S^2.I)$ and $(S^2.II)$ to be optimal, is that there exists an m-component column vector $\lambda \geq 0$ and an n-component row vector $\mu \geq 0$ such that:

(i) $\nabla f(X) = \lambda' G(X) - \mu$

(ii) $\mu X = 0$

(iii) $\lambda' G(X) = 0$

Now $\nabla f(X) = C + 2X'D$

$G(X) = AX - b$

$\nabla G(X) = A$

(See derivations of relations (3.4), (3.5) and (3.6) to obtain these results). Hence if we put $\mu = V'$ we have that a necessary and sufficient condition for a feasible solution X of problems $(S^2_{\cdot}I)$ and $(S^2_{\cdot}II)$ to be optimal, is that there exists an m-component column vector $\lambda \geq 0$ and an n-component column vector $V \geq 0$ such that

$$\left.\begin{array}{lr} \text{(i)}\ 2DX - A'\lambda + V = -C' & (3.10.a) \\ \text{(ii)}\ V'X = 0 & (3.10.b) \\ \text{(iii)}\ \lambda'(AX-b) = 0 & (3.10.c) \end{array}\right] \qquad (3.10)$$

Now if we let $AX-b = -Y$ then $Y \geq 0$ since $AX \leq b$ and (3.10.c) becomes $-\lambda'Y = 0$ or simply

$$\lambda'Y = 0 \qquad (3.11)$$

Hence we can write relations (3.10) in the form

$$2DX - A'\lambda + V = -C'$$

$$V'X + \lambda'Y = 0 \qquad \text{(combining (3.10.b) and (3.11))}$$

$$AX + Y = b$$

where $Y \geq 0$

Whence we alternatively conclude that a necessary and sufficient condition for a feasible solution X of problems $(S^2_{\cdot}I)$ and $(S^2_{\cdot}II)$ to be optimal, is that there exists an m-component column vector $\lambda \geq 0$, an n-component column vector $V \geq 0$ and an m-component column vector $Y \geq 0$ such that

$$\begin{array}{l} 2DX - A'\lambda + V = -C' \\ V'X + \lambda'Y = 0 \\ AX + Y = b. \end{array} \qquad (3.12)$$

Whence X is an optimal solution to Problems $(S^2_{\cdot}I)$ and $(S^2_{\cdot}II)$ if and only if for some λ, Y and V, $[X,\lambda,Y,V]$ satisfies the system

$$\left.\begin{array}{ll} AX + Y = b & (3.13.a) \\ 2DX - A'\lambda + V = -C' & (3.13.b) \\ V'X + \lambda'Y = 0 & (3.13.c) \\ X \geq 0,\ \lambda \geq 0,\ Y \geq 0,\ V \geq 0 & (3.13.d) \end{array}\right] \quad (3.13)$$

This conclusion follows immediately from the statement just preceding relation (3.12), since then X is a feasible solution to Problems $(S^2_{\cdot}I)$ and $(S^2_{\cdot}II)$ ($X \geq 0$ and $AX \leq b$ as $Y \geq 0$) for which there exists $\lambda \geq 0$, $Y \geq 0$, $V \geq 0$ such that $[X,\lambda,Y,V]$ is a solution to system (3.12).

(3.7) <u>An Optimal Solution of a Quadratic Program in Slack Form I or Slack Form II as a Component Vector of a Basic Solution to a System of Equations.</u>

Consider the following quadratic programming problem:

maximise $f(X) = CX + X'DX$

subject to

$AX \leq b$

$X \geq 0$ $\quad (Q^2_{\cdot}S)$

given that the quadratic form $X'DX$ is either negative definite or negative semi-definite.

Problem $(Q^2_{\cdot}S)$ represents a quadratic program either in Slack Form I or a quadratic program in Slack Form II. In section (3.6) (see relation (3.13)) we proved the following result:

<u>Result I.</u> X is a solution to Problem $(Q^2_{\cdot}S)$ if and only if for some λ, Y and V, $[X,\lambda,Y,V]$ satisfies the system

$$\left.\begin{array}{ll} \left.\begin{array}{r} A\,X + Y = \;\; b \\ 2DX - A'\lambda + V = -C' \end{array}\right] & (3.14.a) \\ X \geq 0,\ \lambda \geq 0,\ Y \geq 0,\ V \geq 0 & (3.14.b) \\ V'X + \lambda' Y = 0 & (3.14.c) \end{array}\right] \quad (3.14)$$

We next prove a very important theorem:

<u>THM. IX</u> If Problem (Q^2S) has an optimal solution then there exists at least one basic feasible solution to the system (3.14.a) (i.e., a basic solution to (3.14.a) which satisfies (3.14.b)) satisfying the additional condition

$$V'X + \lambda' Y = 0.$$

PROOF: We give an easy but indirect prrof of this theorem in the end of section (5.4) of Chapter V, when we prove the finite convergence of The Method of Frank and Wolfe. A direct proof of this theorem is extremely difficult to give. Any reader interested in such a proof is referred to the article by H.M. Markowitz published in "Naval Research Logistics Quarterly, Volume 3, 1956". The proof given in section (5.4) is "indirect" in so far as it is a "by product" or rather a conclusion which we reach when proving the finite convergence of The Method of Frank and Wolfe.

REMARK: THM. IX asserts that if Problem (Q^2S) has an optimal solution then there exists at least one basic solution to (3.14.a) say, $[X_o, \lambda_o, Y_o, V_o]$ which also satisfies (3.14.b) and (3.14.c). Hence by Result I it follows immediately that X_o is an optimal solution to Problem (Q^2S). This is a rather important result because then, provided we know that Problem (Q^2S) has an optimal solution, it is only necessary to examine the basic solutions of

(3.14.a) to find a solution satisfying (3.14), i.e., it is only necessary to examine the basic solutions of (3.14.a) to find an optimal solution to Problem ($Q^2_{\cdot}S$). The Method of Frank and Wolfe presented in Chapter V is based upon this important observation.

THM.IX will be used in our next section to prove another important theorem. In order to avoid "confusing notation" we prefer to use $\gamma = [\gamma_1, \ldots, \gamma_m]$ to denote the m component column vector $\lambda = [\lambda_1, \ldots, \lambda_m]$, and restate it in the form:

THM.IX* If Problem ($Q^2_{\cdot}S$) has an optimal solution then there exists at least one basic solution to the system

$$A X + Y = b$$

$$2DX - A'\gamma + V = -C'$$

which satisfies the additional conditions

$$X \geq 0,\ \gamma \geq 0,\ Y \geq 0,\ V \geq 0$$

$$V'X + \gamma'Y = 0.$$

(3.8) An Optimal Solution of a Quadratic Program in Normal Form I or Normal Form II as a Component Vector of a Basic Solution to a System of Equations.

Consider the following quadratic programming problem:

$$\left.\begin{array}{l} \text{maximise} \quad f(X) = CX + X'DX \\ \text{subject to} \\ \quad AX = b \\ \quad X \geq 0\,, \\ \text{given that the quadratic form } X'DX \text{ is either negative definite} \\ \text{or negative semi-definite} \end{array}\right] \quad (Q^2_{\cdot}N)$$

Problem $(Q^2_\cdot N)$ represents a quadratic program either in Normal Form I or a quadratic program in Normal Form II. In section (3.5) (see relation (3.10)) we proved the following result:

Result II. X is a solution to Problem $(Q^2_\cdot S)$ if and only if for some λ, and V, $[X,\lambda,V]$ satisfies the system

$$\left.\begin{array}{ll}\left.\begin{array}{r}AX = b \\ 2DX - A'\lambda + V = -C'\end{array}\right] & (3.15.a) \\ X \geq 0,\ V \geq 0,\ \lambda \ \text{unrestricted} & (3.15.b) \\ V'X = 0 & (3.15.c)\end{array}\right] \quad (3.15)$$

We next prove a very important theorem:

THM. X. If Problem $(Q^2_\cdot N)$ has an optimal solution then there exists at least one basic feasible solution to the system (3.15.a) (i.e., a basic solution to (3.15.a) which satisfies (3.15.b)) satisfying the additional condition

$$V'X = 0$$

PROOF: Write Problem $(Q^2_\cdot N)$ in the equivalent form:

$$\left.\begin{array}{l}\text{maximise} \quad f(X) = CX \\ \text{subject to} \\ \quad AX \leq b \\ \quad -AX \leq -b \\ \quad X \geq 0\end{array}\right] \quad (3.16)$$

Now Problem (3.16) has an optimal solution, therefore, by THM.* IX , it follows that there exists at least one basic solution, say $[X_o, Y^1_o, Y^2_o, Y^1_o, Y^2_o, V_o]$, to the system

$$\left.\begin{array}{llr} AX + Y^1 = b & & (3.17.a) \\ -AX + Y^2 = -b & & (3.17.b) \\ 2DX - A'\gamma^1 + A\gamma^2 + V = -C' & & (3.17.c) \end{array}\right] \qquad (3.17)$$

which satisfies the additional conditions

$$X \geq 0,\ \gamma^1 \geq 0,\ \gamma^2 \geq 0,\ Y^1 \geq 0,\ Y^2 \geq 0,\ V \geq 0 \qquad (3.18)$$

$$V'X + (\gamma^1)'Y_1 + (\gamma^2)'Y_2 = 0 \qquad (3.19)$$

Here we have partitioned the $2m$ component column vectors Y and γ, respectively, into

$$Y = [Y^1, Y^2]$$

$$\gamma = [\gamma^1, \gamma^2]$$

where: Y^i $(i = 1,2)$ is an m-component column vector.

γ^i $(i = 1,2)$ is an m-component column vector.

Now from (3.17.a) and (3.17.b) we obtain that $Y_o^1 + Y_o^2 = 0$ and so $Y_o^1 = 0$ and $Y_o^2 = 0$ since $Y_o^1 \geq 0$ and $Y_o^2 \geq 0$ by (3.18). Thus from (3.19) we obtain that

$$V_o'X_o = 0 \qquad (3.20)$$

Also from (3.17.a) and (3.17.b) we obtain that

$$AX_o = b \qquad (3.21)$$

since $Y_o^1 = Y_o^2 = 0$.

Now if we let $\lambda_o = (\gamma_o^1 - \gamma_o^2)$, then λ_o denotes an m-component column vector of arbitrary sign (since $\gamma_o^1 \geq 0$, $\gamma_o^2 \geq 0$) and from (3.17.c) we have that

$$2DX_o - A'\lambda_o + V_o = -C' \ . \qquad (3.22)$$

Whence since $X_o \geq 0$ and $V_o \geq 0$ it follows from (3.20), (3.21) and (3.22) that $[X_o, \lambda_o, V_o]$ is a solution to the system

$$\begin{bmatrix} A & 0 & 0 \\ 2D & -A' & I \end{bmatrix} \begin{bmatrix} X \\ \lambda \\ V \end{bmatrix} = \begin{bmatrix} b \\ -C' \end{bmatrix} \qquad (3.23)$$

which also satisfies the conditions

$$X \geq 0, \; V \geq 0, \; \lambda \text{ unrestricted}$$

$$V'X = 0 \; .$$

We must still show that $[X_o, \lambda_o, V_o]$ is a basic solution to system (3.23). Now $[X_o, \lambda_o, V_o]$ does not have more than $(m+n)$ components different from zero. (If k components of X_o are different from zero then no more than $(n-k)$ components of V_o are different from zero since $V_o'X_o = 0$. Thus no more than n components of the 2n-component vector $[X_o, V_o]$ can be different from zero. However λ_o contains only m components, and, therefore no more than $(n+m)$ components of $[X_o, \lambda_o, V_o]$ can be different from zero). And so to prove that $[X_o, \lambda_o, V_o]$ is a basic solution to system (3.23), it is sufficient to show that columns of the coefficient matrix

$$M = \begin{bmatrix} A & 0 & 0 \\ 2D & -A' & I_n \end{bmatrix} ,$$

associated in system (3.23), with the non-zero components of $[X_o, \lambda_o, V_o]$, are linearly independent. This follows immediately since the columns of the coefficient matrix

$$N = \begin{bmatrix} A & I & 0 & 0 \\ -A & I_m & 0 & 0 \\ 2D & -A & A & I_n \end{bmatrix}$$

associated in system (3.17), with the non-zero components of the basic solution $[X_o, Y_o^1, Y_o^2, Y_o^1, Y_o^2, V_o]$ are linearly independent, and since

$$Y_o^1 = Y_o^2 = 0$$

$$\lambda_o = (Y_o^1 - Y_o^2) \; .$$

Q.E.D.

REMARK: THM. X asserts that if Problem ($Q^2_\circ N$) has an optimal solution then there exists at least one basic solution to (3.15.a), say, $[X_o, \lambda_o, V_o]$ which also satisfies (3.15.b) and (3.15.c). Hence by Result II it follows immediately that X_o is an optimal solution to Problem ($Q^2_\circ N$). This is rather an important result because then, provided we know that Problem ($Q^2_\circ N$) has an optimal solution, it is only necessary to examine the basic solutions to (3.15.a) to find a solution satisfying (3.15), i.e., <u>it is only necessary to examine the basic solutions to (3.15.a) to find an optimal solution to Problem ($Q^2_\circ N$)</u>. (Wolfe's Method (presented in Chapter IV) and Dantzig's Method (presented in Chapter VI) are both based upon this important observation).

(3.9) <u>Some Useful Properties of Quadratic Forms.</u>

We end Chapter III by deriving some important properties of quadratic forms which are going to be extensively used in the sequel.

<u>THM. XI</u> If $X'DX$ is negative definite (semi-definite) quadratic form, then $Y'D_*Y$ where D_* is a symmetric submatrix of D is also a negative definite (semi-definite) quadratic form.

PROOF: In X, set the variables not associated with the columns of D_* equal to zero. Then the resulting quadratic form must remain negative definite or negative semi-definite (whichever is the case) with respect to the remaining variables, i.e., $Y'D_*Y$ is negative definite (negative semi-definite).

Q.E.D.

THM. XII If the quadratic form $X'DX$ is negative semi-definite then

$DX = 0$ for all X for which $X'DX = 0$.

PROOF: Suppose that $X \neq 0$ and that $X'DX = 0$ then for any Y and arbitrary λ we have

$$0 \geq (Y+\lambda X)'D(Y+\lambda X) = Y'DY + 2\lambda(Y'DX) + \lambda^2(X'DX)$$

$$= Y'DY + 2\lambda(Y'DX) \qquad (3.24)$$

Since λ is arbitrary, relation (3.24) is possible only if the coefficient of λ vanishes, i.e., only if

$$Y'DX = 0 \ .$$

But Y is arbitrary and so we have that,

$$Y'DX = 0 \quad \text{for all } Y.$$

Hence $DX = 0$

Q.E.D.

THM.XIII If $X'DX$ is a negative definite quadratic form then D is non-singular.

PROOF: Since $X'DX$ is negative definite we have that

$$X'DX < 0 \quad \text{for all } X \neq 0 \ ,$$

and so we cannot have $DX = 0$ for $X \neq 0$ otherwise we would obtain a contradiction. Hence the columns of D are linearly independent.

Q.E.D.

REMARKS: (i) The above three simple theorems are all that we need for the purpose of this thesis. Note that they apply equally well to positive quadratic forms.

(ii) If $X'DX$ is a negative definite (semi-definite) the associated matrix D is also called negative definite (semi-definite).

(3.10) Categories of Quadratic Programs.

THM.XIV A quadratic program falls into one of three mutually exclusive and collectively exhaustive categories:

(i) it has no feasible solution

(ii) it has an optimal solution

(iii) it has an unbounded solution, i.e., the objective function is unbounded over the feasible domain.

PROOF: A quadratic program either has a feasible solution or no feasible solution. If it has no feasible solution (i) follows. If it has a feasible solution then:

(a) Either the objective function is bounded above in the feasible domain and hence by THM. VI (and VII) attains its supremum there (which proves (ii)), or

(b) The objective function is not bounded above, in which case we have (iii).

Q.E.D.

REMARK: THM. XIV is easily extended to the case where we have a linear programming problem instead of a quadratic programming problem, since we can then regard the former as a quadratic program where the quadratic objective function $f(X) = CX + X'DX$ has $D = 0$.

PART II

SIMPLICIAL METHODS OF QUADRATIC PROGRAMMING

Chapter IV

"Wolfe's Method"

(4.1) Introduction.

Wolfe's Method is a computational technique developed by Philip Wolfe, for solving quadratic programs in Normal Form I (and II), which uses the simplex algorithm of linear programming with a trivial modification. For reasons to be justified in section (4.5) we present first a shortened version of Wolfe's Method (Short Form), i.e., we deal first with the problem of solving a quadratic program in Normal Form I:

$$\left.\begin{array}{l} \text{maximise} \quad f(X) = CX + X'DX \\ \text{subject to} \\ \qquad AX = b \\ \qquad X \geq 0 , \\ \text{given that the quadratic form } \underline{X'DX \text{ is negative definite}}. \end{array}\right] \qquad (N^2_{\circ}I)$$

The method is based on the following facts (to be proved in Section (4.2)):

(i) There exists an $[X,\lambda,V]$ satisfying the system of equations

$$\begin{bmatrix} A & 0 & 0 \\ 2D & -A' & I_n \end{bmatrix} \begin{bmatrix} X \\ \lambda \\ V \end{bmatrix} = \begin{bmatrix} b \\ -C' \end{bmatrix} \qquad (4.1)$$

and the additional conditions

$$\left.\begin{array}{l} \left.\begin{array}{l} X \geq 0 \\ V \geq 0 \\ \lambda \text{ unrestricted} \end{array}\right] \text{———} (4.2.a) \\ V'X = 0 \qquad\qquad \text{———} (4.2.b) \end{array}\right] \qquad (4.2)$$

if and only if Problem $(N^2_{\circ}I)$ has an optimal solution.

(ii) For any $[X,\lambda,Y]$ satisfying (4.1) and (4.2), X is an optimal solution to Problem ($N^2_\cdot I$).

(iii) If Problem ($N^2_\cdot I$) has an optimal solution there exists at least one basic solution $[X,\lambda,Y]$ to (4.1) satisfying (4.2).

Once the above facts are known we see that the problem of solving the quadratic program ($N^2_\cdot I$) can be reduced to the following: Among the basic solutions of the system (4.1) find one which satisfies (4.2.a) and (4.2.b). (As a matter of fact we shall prove in our next section that, since $X'DX$ is strictly negative definite, there exists only one optimal solution to Problem ($N^2_\cdot I$); see THM. II). Wolfe's Method does exactly this!! Roughly speaking, it consists in constructing an extended system of equations, by introducing "additional" artificial variables in the system (4.1), for which a basic solution and an associated simplex tableau can be found immediately. The additional artificial variables are then driven to zero by means of the simplex method. Through an additional rule for the transition from one basic solution to the next, we are assured that condition (4.2.b) will be satisfied throughout the whole iteration procedure. This additional rule represents the only change from the usual simplex method for linear programming.

Wolfe's Algorithm can be developed in two separate forms, a <u>short</u> and a <u>long</u> form. The long form is composed essentially of two repetitions of the short one, and will be discussed in section 4.5. Note however that while the long form is applicable to a quadratic program in Standard Form II where the quadratic form $X'DX$ is <u>negative semi-definite</u>, the short form yields an optimal point with certainty only if either $C = 0$ or $X'DX$ is negative definite (instead of merely negative semi-definite). We discuss the short-form in section (4.3).

(4.2) Theory of the Method

In Chapter III section (3.6) we proved the following result (see relation (3.9)):

Result I. X is an optimal solution to Problem ($N^2_{\cdot}I$) if and only if for some λ and V, $[X,\lambda,V]$ satisfies

$$AX = b$$

$$2DX - A'\lambda + V = -C'$$

$$V'X = 0$$

$$X \geq 0, \quad V \geq 0, \quad \lambda \text{ unrestricted}$$

(this proves statements (i) and (ii) made in section (4.1)). Hence if Problem ($N^2_{\cdot}I$) has an optimal solution then there exists an $[X,\lambda,V]$ satisfying (4.1), (4.2.a) and (4.2.b). But then by THM. (X.3) there exists at least one basic solution $[X,\lambda,Y]$ to (4.1) satisfying (4.2.a) and (4.2.b) (this proves statement (iii) made in section (4.1)).

Whence we have the following result:

Result II. If Problem ($N^2_{\cdot}I$) has an optimal solution then there exists a basic solution to system (4.1), say, $[X,\lambda,V]$, satisfying (4.2.a) and (4.2.b), whose X-component is an optimal solution to Problem ($N^2_{\cdot}I$).

We next prove two additional results (theorems) needed for a further development in the theory of Wolfe's Method:

THM. I The quadratic program ($N^2_{\cdot}I$) does not have an unbounded solution.

PROOF: We first prove that when $X'DX$ is a strictly negative definite quadratic form, then the function $f(X) = CX + X'DX$ is bounded above over E_n. By THM. (VIII.3) $f(X) \to -\infty$ as $[X] \to \infty$. Hence given any real number M, say, there exists a +ve real number $M_1 > 0$ such that

$$f(X) < M \quad \text{for all } X \text{ such that } [X] > M_1 \tag{4.3}$$

Next consider any point of E_n, say, X_o and let $F(X_o) = CX_o + X_o'DX_o = a$, where "a" is a real number. Then by (4.3), there exists a positive real number, say, M_a such that

$$f(X) < a \quad \text{for all } X \text{ such that } [X] > M_a \tag{4.4}$$

Now the function $f(X)$ attains its supremum in the closed and bounded set $\mathbb{B} = \{X | [X] \leq M_a\}$, since $f(X)$ is continuous on E_n and since any continuous function attains its supremum on a closed and bounded set. This means that there exists a point (vector) $X_* \in \mathbb{B}$ such that

$$f(X) < f(X_*) \quad \text{for all } X \in \mathbb{B} \tag{4.5}$$

Hence from (4.4) and (4.5) we conclude that

$$f(X) < \max\{a, f(X_*)\} \quad \text{for all } X \in E_n ,$$

that is $f(X) = CX + X'DX$ is bounded above on E_n whenever the quadratic form $X'DX$ is strictly negative definite.

Now to prove the theorem we only need to note that the feasible domain $\mathbb{K}_N = \{X | AX = b, X \geq 0\}$, of the quadratic program (N^2I), is a subset of E_n and so the objective function $f(X) = CX + X'DX$ must be bounded above over $\mathbb{K}_N$, otherwise $f(X) = CX + X'DX$ could not be bounded above, over E_n. Whence the quadratic program (N^2I) does not have an unbounded solution.

Q.E.D.

<u>THM. II</u> The quadratic program (N^2I) either has a <u>unique</u> optimal solution or no feasible solution.

PROOF: By THM.(XI.3) we have that any quadratic program falls into one of three mutually exclusive and collectively exhaustive categories:

(i) It has no feasible solution.

(ii) It has an optimal solution.

(iii) It has an unbounded solution.

By THM. I, (iii) is impossible. Hence (N^2I) either has no feasible solution or then it has an optimal solution. To prove the uniqueness of the optimal solution when it exists, we note that, since $X'DX$ is strictly negative definite, the objective function $f(X) = CX + X'DX$ is strictly concave (THM.V.3) and so by THM.$(VIII^*.1)$ $f(X)$ takes on its (global) maximum at a unique point of $\mathbb{K}_N$.

Q.E.D.

Now Problem (N^2I) either has a feasible solution or no feasible solution. If it has a feasible solution then it has a unique optimal solution and so by Result II above there exists a basic solution to system (4.1), say, $[X_o, \lambda_o, Y_o]$, satisfying (4.2.a) and (4.2.b), whose X-component X_o, is the optimal solution of Problem (N^2I). Bearing in mind these facts Wolfe's Method (Short Form) is divided into two Phases, Phase I and Phase II:

<u>Phase I.</u> In this Phase by introducing artificial variables into the system (4.1) and expanding it to an "enlarged" system (see (4.6) below), the constraints $AX = b$ are tested for feasibility. If these constraints are feasible, a basic solution to the "enlarged system", satisfying (4.2.a) and (4.2.b) (but not necessarily (4.1)), is produced with which to start Phase II.

<u>Phase II.</u> In this Phase we drive to zero the artificial variables, present as basic variables in the basic solution obtained at the end of Phase I, over a sequence of basic solutions which satisfy (4.2.a) and (4.2.b). In the end of Phase II we obtain a basic solution to (4.1) satisfying (4.2.a) and (4.2.b) whose X-component is an optimal solution to Problem (N^2I).

(4.3) Wolfe's Algorithm (Short Form)

Phase I. Expand the system (4.1) into the form

$$\left.\begin{array}{l} AX + Y = b \\ 2DX - A'\lambda + V + Z_1 - Z_2 = -C' \\ X \geq 0 \ , \quad V \geq 0, \ \lambda \text{ unrestricted} \\ Y \geq 0 \ , \quad Z_1 \geq 0 \ , \quad Z_2 \geq 0 \end{array}\right| \qquad (4.6)$$

by introducing the following $m + 2n$ non-negative artificial variables

$$Y = [y_1, y_2, \ldots, y_m]$$
$$Z_1 = [z_1^1, z_2^2, \ldots, z_n^1]$$
$$Z_2 = [z_1^2, z_2^2, \ldots, z_n^2]$$

Then, clearly, $[\overline{X}, \overline{\lambda}, \overline{V}, \overline{Z}_1, \overline{Z}_2]$, where:

(a) $\overline{X} = 0, \overline{\lambda} = 0, \overline{V} = 0, \overline{Y} = b$ (≥ 0 by hypothesis)

(b) $\overline{z}_j^1 = -c_j$ and $\overline{z}_j^2 = 0$ if c_j is negative
$\overline{z}_j^1 = 0$ and $\overline{z}_j^2 = c_j$ if c_j is positive
$\overline{z}_j^1 = 0$ and $\overline{z}_j^2 = 0$ if $c_j = 0$

is a basic solution (feasible) to the system of constraints (4.6) satisfying (4.2.a) and (4.2.b), i.e., satisfying (4.2.a) and $V'X = 0$. With this initial basic solution to the system (4.6) use the simplex algorithm to maximise the linear form

$$z_1^* = - \sum_{i=1}^{m} y_i \qquad (4.7) \qquad (L_1)$$

subject to the constraints (4.6) as well as to the constraints $\lambda = 0$ and $V = 0$.

If the constraints of Problem (N^2I) are consistent, the maximum for (4.7) will be zero and we will obtain a basic solution to (4.6) (see App. IV.1), say, $[X,\lambda,V,Y,Z_1,Z_2] = [X_*,0,0,0,Z_1^*,Z_2^*] = \phi_*^*$ which has m of the n variables x_j and n of the $2n$ variables z_j^1, z_j^2 in the basis.

Nevertheless note that the z-variables of ϕ^* obey the following restriction: <u>if z_j^1 is in the basis then z_j^2 is not in the basis and vice-versa.</u> We next prepare the initiation of Phase II in the following way:

(1) In the final simplex tableau of Phase I delete the columns corresponding to all the y_i and to those z_j^1 and z_j^2 not in the basis.

(2) Form an n-component column vector $Z_* = [z_1^*,\ldots,z_n^*]$ from the remaining basic variables z_j^1, z_j^2, where z_j^* is either $(z_j^1)^*$ (a component of Z_1^*) or $(z_j^2)^*$ (a component of Z_*^2), depending on which of the two is left in the basis of the simplex tableau obtained at the end of Phase I.

(3) Form the $(n\times n)$ matrix $F = \|f_{ij}\|$ where

f_{jj} = coefficient of the component z_j^* in Z^* for $j = 1,\ldots,n$

$f_{ij} = 0$ for $i \neq j$

F is then a diagonal matrix with elements $+1$ or -1 depending on whether $z_j^* = (z_j^1)^*$ or $z_j^* = (z_j^2)^*$.

Thus if we let $Z = [z_1,\ldots,z_n]$ be an n-component column vector of artificial variables then $[X,\lambda,Y,Z] = [X_*,0,0,Z_*]$ is a basic ("feasible") solution to the system:

$$\left.\begin{array}{ll}\begin{bmatrix} A & 0 & 0 & 0 \\ 2D & -A' & I_n & F \end{bmatrix} \begin{bmatrix} X \\ \lambda \\ V \\ Z \end{bmatrix} = \begin{bmatrix} b \\ -C' \end{bmatrix} & (4.8.a) \\ X \geq 0,\quad V \geq 0,\quad Z \geq 0,\quad \lambda \text{ unrestricted} & (4.8.b) \end{array}\right] \quad (4.8)$$

<u>Hence at the end of Phase I we will have a basic (feasible) solution with an associated simplex tableau</u> for the system (4.8). The existence of this basic solution implies that the system $AX = b$ is feasible ($X_* \geq 0$ is a solution to $AX = b$) and whence we immediately know:

(i) that Problem (N^2.I) has an optimal solution (THM. II) and

(ii) that there exists a basic solution (unique) to (4.1) satisfying (4.2.a) and (4.2.b), whose X component is an optimal solution to Problem N^2.I (Result II).

With the basic feasible solution $[X_*, 0, 0, Z_*]$ we are ready to initiate Phase II, but before we do so note the following two remarks:

REMARKS: (i) In the likely event of degeneracy we must be sure to continue with Phase I until no y_i appears in the basis, even with the value zero (i.e. even at zero level). It is not sufficient to stop Phase I as soon as we obtain $z_1^* = 0$, for then if there exists an artificial variable y_i at zero level in the basis, we do not immediately have a basic feasible solution to $AX = b$, and this is what we are after.

(ii) In all of Phase I the λ_i and v_i stay out of the basis.

Phase II. In this Phase the Simplex Method is used to solve the following linear programming problem:

$$\text{maximize} \quad z_2^* = -\sum_{j=1}^{n} z_j$$

subject to the constraints (4.8), i.e., subject to

$$\begin{bmatrix} A & 0 & 0 & 0 \\ 2D & -A' & I_n & F \end{bmatrix} \begin{bmatrix} X \\ \lambda \\ V \\ Z \end{bmatrix} = \begin{bmatrix} b \\ -C' \end{bmatrix} \qquad (4.9) \qquad (L_2)$$

$X \geq 0$, $V \geq 0$, $Z \geq 0$

λ unrestricted.

But since in problem (L_2) the components of λ are unrestricted in sign we use the usual transformation $\lambda = (\omega - \eta)$, $\omega \geq 0$, $\eta \geq 0$, to obtain all the variables positive and then apply the Simplex Method, instead, to the equivalent problem

$$\text{maximize} \quad z_2^* = -\sum_{j=1}^{n} z_j$$

subject to the constraints

$$\begin{bmatrix} A & 0 & 0 & 0 & 0 \\ 2D & -A' & A' & I_n & F \end{bmatrix} \begin{bmatrix} X \\ \omega \\ \eta \\ V \\ Z \end{bmatrix} = \begin{bmatrix} b \\ -C' \end{bmatrix} \qquad (4.10) \qquad (L_3)$$

$$X \geq 0,\ \omega \geq 0,\ \eta \geq 0,\ V \geq 0,\ Z \geq 0$$

Note, however, that in the application of the simplex algorithm to Problem (L_3) we introduce one variation from the standard simplex procedure, namely, an additional rule which is to be observed for all indices j and all transitions from one simplex tableau to the next:

<u>Additional Rule</u>: Given a basic feasible to (4.10), if x_j is in the basis (i.e., if x_j is a basic variable), then in the transition to the next basic solution, v_j may not be taken into the basis; if v_j is in the basis, then in the transition to the next basic solution, x_j is not allowed to enter the basis.

This rule ensures that x_j and v_j will not be in the basis at any one iteration of the simplex iteration. Thus $V'X = 0$ for all basic feasible solutions dealt with in Phase II and so condition (4.2.b) will be satisfied during the whole procedure.

Stop Phase II, as soon as, the objective function $z_2^* = 0$, for then we have found a solution to (4.10), i.e., to (4.9) with $Z = 0$ and, with it, a solution to (4.1), which also satisfies conditions (4.2.a) and (4.2.b). Therefore, we will have an optimal solution for the original Problem (N^2I). In the next section we shall prove that Phase II always comes to a stop, i.e., z_2^* will equal zero at some iteration step of Phase II.

(4.4) Finite Convergence of Wolfe's Algorithm.

In this section we shall prove that in the absence of degeneracy in the system (4.10), at some iteration step of Phase II, $z_2^* = 0$ for some basic feasible solution of Problem (L_3), provided D is a negative definite matrix, i.e., we shall prove that when the quadratic program is in Normal Form I then Phase II always terminates with an optimal solution to the program. Note that the assumption of non-degeneracy is not restrictive, however, since any system of equations is treated as if it were not degenerate by the Simplex Method (see App. IV.2).

Proof of Convergence: Since the objective function z_2^* of Problem (L_3) is bounded above by 0 we have that Problem (L_3) either has no feasible solution or an optimal solution (see App. IV.3). The former case is clearly impossible since we are in Phase II and hence have initiated it with a basic feasible solution to Problem (L_3). Hence Problem (L_3) has an optimal solution and $\max z_2^* \leq 0$ for all feasible solutions of Problem (L_3). Now since the system of constraints of Problem (L_3) is non-degenerate there can be no "cycling", and so when applying the simplex algorithm to Problem (L_3) we will:

(1) Either obtain a "simplex tableau" with all the quantities $Z_j - h_j \geq 0$, where Z_j is the "marginal substitution coefficient" of Problem (L_3) (see App. IV.4) and h_j is the "price" associated with each variable of Problem (L_3) ($h_j = 0$ for the "legitimate" variables and $h_j = -1$ for the artificial variables z_j). In other words, we will either obtain an optimal solution to Problem (L_3) or

(2) Reach an iteration where, due to the Additional Rule, no further iterations can be made without violating the simplex rules, that is, we will reach an iteration where we cannot make additional iterations which do not violate the simplex rules or the condition $V'X = 0$.

Now, for any optimal solution of Problem (L_3) we have $z_2^* = 0$, i.e., max z_2^* is equal to zero over the feasible domain of Problem (L_3). This is justified in the following way:

Since we are in Phase II the system $AX = b$ is feasible and so by THM. II Problem (N^2.I) has an optimal solution. Hence by Result II there exists a basic solution to system (4.1) satisfying (4.2.a) (and (4.2.b), for that matter; however we do not need this extra condition for the present argument). But any basic solution satisfying (4.1) and (4.2.a) is a solution (feasible) to Problem L_2 and hence to Problem L_3, with $Z = 0$. Therefore for this solution we have $z_2^* = 0$. But $z_2^* \leq 0$ for all feasible solutions to Problem (L_3). Whence max z_2^* is equal to zero over the feasible domain of Problem (L_3), i.e., for any optimal solution of Problem (L_3) we have $z_2^* = 0$. Thus if case (1) above happens, we have reached an iteration where $z_2^* = 0$ for some basic feasible solution of Problem (L_3) i.e., we have found a basic solution to 4.10 (and 4.9) with $Z = 0$, i.e., a solution to (4.1), which also satisfies (4.2.a) and (4.2.b).

Consequently we will have found an optimal solution to the original problem $N^2_{\cdot}I$, namely, the X-component of the above basic solution.

<u>If case (2) above happens</u>, then we will also have reached an iteration, where $z_2^* = 0$ for some basic feasible solution to 4.10 (and 4.9), i.e., we will also have found a basic solution to 4.10 (and 4.9) with $Z = 0$. To prove this we need Lemma I below.

Before proving Lemma I we need the following "set up":

(a) Let W be a vector of ℓ variables, h a row vector with ℓ constant components, G an $n\times m$ matrix, and g a column vector with n constant components.

(b) Let the symbols $A, C, D, X, \omega, \eta, V$ and b have the same meaning as in section (4.3) with the exception that here we are allowing the matrix D to be either negative semi-definite or strictly negative definite, instead of just negative definite.

(c) Let $\hat{U} = [\hat{X}, \hat{\omega}, \hat{\eta}, \hat{V}, \hat{W}] \geq 0$ be a column vector (of constants) such that $\hat{V}'\hat{X} = 0$. And for this column vector $\hat{U}$ let:

(c_1) $\hat{X}_1$ contain the components of $\hat{X}$ which are positive, and $\hat{V}_1$ contain the corresponding components of $\hat{V}$ (note that then $\hat{V}_1 = 0$ since $\hat{V}_1'\hat{X}_1 = 0$).

(c_2) $\hat{V}_2$ contain the components of $\hat{V}$ which are positive, and $\hat{X}_2$ contain the corresponding components of $\hat{X}$. (Note that then $\hat{X}_2 = 0$ since $V_2'X_2 = 0$).

(c_3) $\hat{X}_3$ contain the remaining components of $\hat{X}$ which are zero, i.e., let $\hat{X}_3$ contain those components of $\hat{X}$ which are zero, but not required to be zero because the corresponding components of $\hat{V}$ are zero. Similarly let $\hat{V}_3$ contain the remaining components of $\hat{V}$, i.e., let $\hat{V}_3$ contain those components of $\hat{V}$ which are zero, but are not required to be zero because the corresponding components of $\hat{X}$ are zero.

LEMMA I. Suppose that $\hat{U} = [\hat{X},\hat{\omega},\hat{\eta},\hat{V},\hat{W}]$ is an optimal solution to the following linear programming problem.

$$\left.\begin{array}{l}
\text{maximise} \quad z = hW \\
\text{subject to the constraints} \\
\qquad AX = b \\
2DX - A'\omega + A'\eta + V + GW = g \\
X \geq 0, \omega \geq 0, \eta \geq 0, V \geq 0, W \geq 0 \\
X_2 = 0, V_1 = 0
\end{array}\right| \qquad (L)$$

then:

(i) If the matrix D is negative semi-definite there exists an n component row vector $\hat{t}$, say, such that

$$h\hat{W} = \hat{t}g$$

(ii) If the matrix D is strictly negative definite then $\hat{t} = 0$ and so $h\hat{W} = 0$.

PROOF: If we let $\lambda = [\omega - \eta]$ then Problem (L) becomes

$$\left.\begin{array}{l}
\text{maximise} \quad z = hW \\
\text{subject to} \\
\qquad AX = b \\
2DX - A'\lambda + V + GW = g \\
X \geq 0, V \geq 0, W \geq 0, \lambda \text{ unrestricted} \\
X_2 = 0, V_1 = 0 .
\end{array}\right| \qquad (L^*)$$

Now, after a possible reordering of indices, partition the vectors X and V as follows:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \qquad V = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix},$$

and partition the matrices A, D, G, I and the vector g, correspondingly, as follows:

$$A = (A_1 \;\; A_2 \;\; A_3)$$

$$D = \begin{bmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{bmatrix} \quad I = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{bmatrix} \quad G = \begin{bmatrix} G_1 \\ G_2 \\ G_3 \end{bmatrix} \quad g = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}$$

Then Problem (L^*) can be written in the form:

$$\text{maximise} \quad z = hW$$

subject to the constraints

$$\left.\begin{array}{l} A_1X_1 + A_2X_2 + A_3X_3 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 = b \\ 2D_{11}X_1 + 2D_{12}X_2 + 2D_{13}X_3 - A_1'\lambda + I_1V_1 + G_1W = g_1 \\ 2D_{21}X_1 + 2D_{22}X_2 + 2D_{23}X_3 - A_2'\lambda + I_2V_2 + G_2W = g_2 \\ 2D_{31}X_1 + 2D_{32}X_2 + 2D_{33}X_3 - A_3'\lambda + I_3V_3 + G_3W = g_3 \end{array}\right] (4.11)$$

$X_1 \geq 0,\ X_3 \geq 0,\ V_2 \geq 0,\ V_3 \geq 0,\ W \geq 0$

λ unrestricted

$X_2 = 0,\ V_1 = 0$. (L-P)

Since $X_2 = 0$ and $V_1 = 0$ the dual of the primal problem (L-P) can be written in the form (check with the definition of the dual)

$$\left.\begin{array}{l} \text{minimise } Z = sb + t_1g_1 + t_2g_2 + t_3g_3 \\ \text{subject to} \\ \left.\begin{array}{llr} sA_1 + 2t_1D_{11} + 2t_2D_{21} + 2t_3D_{31} \geq 0 & & (4.12.a) \\ sA_3 + 2t_1D_{13} + 2t_2D_{23} + 2t_3D_{33} \geq 0 & & (4.12.b) \\ t_1A_1' + t_2A_2' + t_3A_3' = 0 & & (4.12.c) \\ t_2 \geq 0 & & (4.12.d) \\ t_3 \geq 0 & & (4.12.e) \\ t_1G_1 + t_2G_2 + t_3G_3 \geq h' & & (4.12.f) \end{array}\right] (4.12) \end{array}\right] \text{(L-D}$$

where s is an m-component row vector (of variables) and $t = (t_1,t_2,t_3)$ is an n-component row vector (of variables) and where there is <u>no further sign restriction on the variables of s and t</u>.

<u>Note</u>: (1) that the (m+n)-component vector (s,t_1,t_2,t_3) has been partitioned in the same way as the rows of (4.11)

(2) that since the components of λ are unrestricted in sign the corresponding dual constraints i.e., (4.12.c), must hold as strict equalities.

Now by the fundamental theorem of duality (see App. IV.5) since the primal problem (L-P) has an optimal solution namely, $[\hat{X},\hat{\lambda},\hat{V},\hat{W}] = [\hat{X}_1,\hat{X}_2,\hat{X}_3;\hat{\lambda};\hat{V}_1,\hat{V}_2,\hat{V}_3;\hat{W}]$, the dual problem (L-P) also has an optimal solution, say, $[\hat{s},\hat{t}_1,\hat{t}_2,\hat{t}_3]$ and

$$\max z = h\hat{W} = \hat{s}b + \hat{t}_1g_1 + \hat{t}_2g_2 + \hat{t}_3g_3 = \min Z. \qquad (4.13)$$

Furthermore, by the complementary slackness property (see App.IV.6), for those signs restricted variables (≥ 0) of Problem (L-P) which <u>are positive</u> in the optimal solution $[\hat{X}_1,\hat{X}_2,\hat{X}_3;\ \hat{\lambda};\hat{V}_1,\hat{V}_2,\hat{V}_3,\hat{W}]$, namely $\hat{X}_1$ and $\hat{V}_2$, the corresponding inequalities of Problem (L-D) (that is, (4.12.a) and (4.12.d)) are satisfied as strict equalities for the optimal solution

$\hat{s},\hat{t}_1,\hat{t}_2,\hat{t}_3$. And so we have that

$$\left.\begin{array}{lr} \hat{s}A_1 + 2\hat{t}_1D_{11} + 2\hat{t}_2D_{21} + 2\hat{t}_3D_{31} = 0 & (4.14.a) \\ \hat{s}A_3 + 2\hat{t}_1D_{13} + 2\hat{t}_2D_{23} + 2\hat{t}_3D_{33} \geq 0 & (4.14.b) \\ \hat{t}_1A_1' + \hat{t}_2A_2' + \hat{t}_3A_3' = 0 & (4.14.c) \\ \hat{t}_2 = 0 & (4.14.d) \\ \hat{t}_3 \geq 0 & (4.14.e) \\ \hat{t}_1G_1 + \hat{t}_2G_2 + \hat{t}_3G_3 \geq h' & (4.14.f) \end{array}\right] \quad (4.14)$$

Postmultiplying (4.14.a) and (4.14.b) by $\hat{t}_1'$ and $\hat{t}_3'$ respectively we obtain

$$\hat{s}A_1\hat{t}_1' + 2\hat{t}_1D_{11}\hat{t}_1' + 2\hat{t}_3D_{31}\hat{t}_1' = 0 \qquad \text{since } \hat{t}_2 = 0 \tag{4.15}$$

$$\hat{s}A_3\hat{t}_3' + 2\hat{t}_1D_{13}\hat{t}_3' + 2\hat{t}_3D_{33}\hat{t}_3' \geq 0 \qquad \text{since } \hat{t}_2 = 0 \text{ and } \hat{t}_3 \geq 0 \tag{4.16}$$

Adding relations (4.15) and (4.16) we get

$$\underbrace{\hat{s}\left[A_1\hat{t}_1 + A_3\hat{t}_3\right]}_{\substack{\shortparallel \\ M}} + \underbrace{2(\hat{t}_1,\hat{t}_3)\begin{bmatrix} D_{11} & D_{13} \\ D_{31} & D_{33} \end{bmatrix}\begin{bmatrix} \hat{t}_1' \\ \hat{t}_3' \end{bmatrix}}_{\substack{\shortparallel \\ N}} \geq 0 \tag{4.17}$$

But from (4.14.c) and (4.14.d) it follows that $M = 0$.

And so we have that

$$N \geq 0 \tag{4.18}$$

At this stage we break up our analysis into the two cases, (i) and (ii) ((i) will prove statement i of our lemma and (ii) will prove statement ii):

(i) <u>If D is a negative semi-definite matrix</u>, then

$$D_* = \begin{bmatrix} D_{11} & D_{13} \\ D_{31} & D_{33} \end{bmatrix}$$

is a negative semi-definite matrix. This follows since D is a symmetric submatrix of D (by THM. XI.3). Therefore $N \leq 0$. Hence from (4.18) we conclude that $N = 0$, i.e.,

$$2(\hat{t}_1,\hat{t}_3)\begin{bmatrix} D_{11} & D_{13} \\ D_{31} & D_{33} \end{bmatrix}\begin{bmatrix} \hat{t}_1' \\ \hat{t}_3' \end{bmatrix} = 0 \tag{4.19}$$

Hence from THM.(XII.3) we have that

$$\begin{bmatrix} D_{11} & D_{13} \\ D_{31} & D_{33} \end{bmatrix}\begin{bmatrix} \hat{t}_1' \\ \hat{t}_3' \end{bmatrix} = 0 \tag{4.20}$$

that is

$$D_{11}\hat{t}_1' + D_{13}\hat{t}_3' = 0 \tag{4.21}$$

$$D_{31}\hat{t}_1' + D_{33}\hat{t}_3' = 0 \tag{4.22}$$

If we substitute equations (4.21) and (4.14.d) in relation (4.14.a) we obtain

$$\hat{s}A_1 = 0 \tag{4.23}$$

and therefore

$$\hat{s}b = \hat{s}(A_1\hat{X}_1 + A_2\hat{X}_2 + A_3\hat{X}_3) \quad \text{from (4.11)} \tag{4.24}$$

$$= 0 ,$$

since $\hat{X}_2 = 0$ and $\hat{X}_3 = 0$.

Using (4.24) in relation (4.13) we obtain

$$h\hat{W} = \hat{t}_1 g_1 + \hat{t}_2 g_2 + \hat{t}_3 g_3 = \hat{t}g , \tag{4.25}$$

which proves statement (i) of our lemma.

(ii) <u>If D is a strictly negative definite matrix</u> then D is a negative semi-definite matrix and so we can make use of relations (4.19) and (4.20) derived for case (i), but now the matrix D_* is <u>strictly negative</u> definite and so relation (4.20) implies that $(\hat{t}_1, \hat{t}_3) = 0$ (see THM.XIII.3). Hence $\hat{t}_1 = 0$, $\hat{t}_2 = 0$ (from 4.14.d), and $\hat{t}_3 = 0$, whence

$$\hat{t} = (\hat{t}_1,\hat{t}_2,\hat{t}_3) = 0$$

and $\quad h\hat{W} = 0 \quad$ from (4.13) .

This proves statement (ii) of our lemma.

Q.E.D.

Now the basic feasible solution $U_o = [X_o, \omega_o, \eta_o V_o Z_o]$, say, which is obtained in case (2) of Phase II satisfies $V_o' X_o = 0$, since we have been observing the Additional Rule at every iteration of Phase II. And so if we let:

(1) X_1^o contain the components of X_o which are positive, and V_1^o contain the corresponding components of V_o (note that then $V_1^o = 0$ since $(V_1^o)' X^o = 0$).

(2) V_2^o contain the components of V_o which are positive, and X_2^o contain the corresponding components of X_o (note that then $X_2^o = 0$ since $(V_2^o)' X_2^o = 0$).

(3) X_3^o contain those components of X_o which are zero, but are not required to be zero because the corresponding components of V_o, denoted by V_3^o, are zero.

(4) $W = Z$ with $\ell = n$ (remember W was assumed to have ℓ components)

$h = (-1, \ldots, -1)$

$G = F$

$g = -C'$

then $V_o = [X_o, \omega_o, \eta_o, V_o, Z_o]$ will be an optimal solution to the following linear programming problem (compare with Problem (L_2)):

$$\text{maximise } z = hZ = -\sum_{j=1}^{n} z_j$$

subject to the constraints

$$\left.\begin{aligned} AX &= b \\ 2DX - A'\omega + A'\eta + V + FZ &= -C' \\ X \geq 0,\ \omega \geq 0,\ \eta &\geq 0,\ V \geq 0, Z \geq 0 \\ X_2 = 0,\ V_1 &= 0 \end{aligned}\right] \quad (4.26)$$

(L_4)

where: (i) the components of X_2 are the x-variables comprising X_2^o (i.e., where the components of X_2 are the variables of X, with the value zero of course, which form X_2^o)

and where the components of V_1 are the v-variables comprising V_1^o (i.e., where the components of V_1 are the variables of V, with the value zero of course, which form V_1^o.

(ii) D is strictly negative definite. L_4

Justification of the Optimality of U_o: The basic feasible solution $U_o = [X_o, \omega_o, \eta_o, V_o, Z_o]$ which is obtained in case (2) of Phase II, is an optimal solution to Problem (L_4), because case (2) of Phase II is the case, if you recall, where we obtain a simplex tableau of Problem (L_3) for which we cannot make additional iterations without violating the Simplex Rules or the Additional Rule $V'X = 0$. This is equivalent to making the following statement:

Statement A. In case (2) of Phase II we are in presence of a simplex tableau to Problem (L_3) (and in presence of a basic feasible solution to Problem (L_3) namely U_o) in which all the $Z_j - h_j \geq 0$ (the price h_j of all "legitimate" variables is zero and the price h_j of all the "artificial" variables z_j is -1) for all the variables not contained in V_1 or X_2, where, as before, Z_j is the marginal substitution coefficient of Problem (L_3).

But $U_o = [X_o, \omega_o, \eta_o, V_o, Z_o]$ is also a basic feasible solution to Problem (L_4) since it satisfies (4.26) and hence is an optimal solution to Problem (L_4) by Statement (A) above.

REMARK: Note that we are assuming here that the system of constraints of Problem (L_3) is not degenerate so that all basic variables of U_o are different from zero.

Hence we finally arrive at the following conclusion:

Conclusion A. If we now use LEMMA I with U_o and Problem (L_4) as underlying hypothesis we obtain, from Statement (ii) of this Lemma, that

$$-hZ_o = - \sum_{j=1}^{n} z_j^o = 0$$

i.e., we obtain that all the artificial variables z_j present in the basic feasible solution $U_o = [X_o, \omega_o, \eta_o, V_o, Z_o]$ of Problem (L_3) are equal to zero. So we have managed to prove that the basic feasible solution $U_o = [X_o, \omega_o, \eta_o, V_o, Z_o]$ to the systems (4.10) and (4.9) which we obtain in case (2) of Phase II has $Z_o = 0$.

This is, if you recall, what we wanted to show!! Thus Wolfe's Method - Short Form definitely leads to an optimal solution of a quadratic program in Normal Form I, i.e., to a quadratic problem where the form $X'DX$ is strictly negative definite.

REMARKS: (i) It should be noted that the Short Form leads definitely also to an optimal solution of quadratic programs in Normal Form II (where the quadratic form $X'DX$ is negative semi-definite) provided the objective function $f(X) = CX + X'DX$ of these programs has no linear term, i.e., provided $C = 0$. This follows immediately by using statement (i) of Lemma I in the Conclusion A above, since then $hZ_o = -\hat{t}C' = 0$.

(ii) It should be noted that the Short Form does not necessarily lead (at least in theory) to an optimal solution of a quadratic programs in Normal Form II (where the quadratic form $X'DX$ is negative semi-definite) if the objective function $f(X) = CX+X'DX$ of these programs have $C \neq 0$, because then we do not necessarily have $hZ_o = -\hat{t}C' = 0$.

(4.5) Wolfe's Algorithm (LONG FORM).

The Long Form is a computational technique also developed by Philip Wolfe, for solving quadratic programs in Normal Form II where the quadratic form $X'DX$ is negative semi-definite (instead of strictly negative definite as in the Short Form), and where the objective function has a linear term (i.e. where $C \neq 0$). It is composed essentially of two repetitions of the Short Form. While the elegant theory of the LONG Form is a thing to marvel at (see App. IV.4) we find that its inclusion, in a Computer Program designed for the Short Form, is not a procedure that "pays off" as far as practical applications go, for it can become extremely time consuming in quadratic problems of a fairly large size. So whenever we are met with a quadratic program in Normal Form II we prefer either:

(1) to apply Dantzig's Method (see Chapter VI) or Beale's Method (see Chapter VII); or

(2) to make use of the "Razor's Edge Property" of negative semi-definite quadratic forms, suggested by E.M.L. Beale. By the "Razor's Edge Property" we mean that a negative semi-definite form $X'DX$ can be converted into a negative (strictly) definite form by making an arbitraly small change in the D. More specifically (and here we quote from (4) in the bibliography:- "If $X'DX$ is negative semi-definite, then $X'(D+\varepsilon I)X$ is negative definite for any $\varepsilon < 0$, however small $[\varepsilon]$. To prove this, note that $X'DX \leq 0$ for any X, and $\varepsilon X'IX < 0$ for any $X \neq 0$. Thus $X'(D+\varepsilon I)X < 0$ for any $X \neq 0$, and the form is negative definite. Consequently, if we have a quadratic programming problem in which we know that $X'DX$ is either negative

definite or negative semi-definite, we can make sure that the form is negative definite by subtracting a unit in perhaps the fourth of fifth decimal place of each diagonal element of D. In this fashion, the perturbation can be made small enough that in no way affects the numerical results obtained. The one case where such a perturbation could influence the numerical answer is that where $X'DX$ was originally indefinite and the problem had an unbounded solution. However, just as in linear programming, properly formulated problems should not have unbounded solutions, and hence no difficulty should arise".

Generally, Wolfe's Method - Short Form derived in section (4.3) will work even for quadratic programs in Normal Form II (i.e., even when $X'DX$ is negative semi-definite), and consequently the usual procedure is to try it without attempting first to perturb D.

(4.6) A Flow of Wolfe's Algorithm (Short Form) in Summary Form:

(a) Given a quadratic program in Normal Form I, i.e.,

$$\text{maximise} \quad f(X) = CX + X'DX \quad (D \text{ strictly negative definite})$$

subject to

$$AX = b$$

$$X \geq 0 \ ,$$

form the system

$$AX + Y = b \qquad (4.27)$$

$$2DX - A'\lambda + V + Z_1 - Z_2 = -C'$$

where:

$$Y = [y_1, y_2, \ldots, y_m]$$
$$Z_1 = [z_1^1, z_2^1, \ldots\ldots\ldots, z_n^1]$$
$$Z_2 = [z_1^2, z_2^2, \ldots\ldots\ldots, z_n^2]$$

are (m+2n) artificial variables. Initiate Phase I.

(b) Phase I:

Step I. Use the Simplex Method to maximise

$$z_1^* = -\sum_{i=1}^{m} y_i$$

to zero, subject to the constraints (4.27), and to the constraints

$$X \geq 0,\ Y \geq 0,\ Z_1 \geq 0,\ Z_2 \geq 0,\ \lambda = 0,\ V = 0$$

If max $z_1^* < 0$, then the system $AX = b$ is not feasible and the Problem (N^2I) has no feasible solution.

If max $z_1^* = 0$, then the system $AX = b$ is feasible and therefore it has an optimal solution (THM.II). Perform Step II.

Step II. Discard Y and the nonbasic components of Z_1 and Z_2. Let the remaining n components be denoted by $Z = [z_1, \ldots, z_n]$ and let their coefficients be the elements (diagonal) of a diagonal matrix denoted by F. Form the system:

$$\begin{aligned} AX &= b \\ 2DX - A'\omega + A\eta + I_n V + FZ &= -C' \end{aligned} \qquad (4.28)$$

Perform Phase II.

(c) Phase II. Use the Simplex to maximise

$$z_2^* = -\sum_{j=1}^{n} z_j$$

subject to the constraints (4.28),

$$X \geq 0,\ \omega \geq 0,\ \eta \geq 0,\ V \geq 0,\ Z \geq 0$$

and to the following side condition:- For $j = 1,\ldots,n$, if x_j is in the basis, we do not admit v_j into the basis; if v_j is in the basis, we do not admit x_j into the basis. The X-component of the terminal basic feasible solution is an optimal solution to the initial Problem ($N^2_\circ I$).

Computation Time. It will take at most $\binom{3n}{n}$ iterations to reach an optimal solution of Problem ($N^2_\circ I$). The important thing to notice on Wolfe's Method is that for quadratic problems of small size it requires a relatively "big" number of simplex iterations to reach an optimal solution. And this is a factor to reckon with if the small problem in question is not being solved by a computer. For these occasions Beale's Method seems to be the best answer. With "large" size quadratic problems the above factor does not enter into consideration for then the problem must be solved in a computer and any computer program for the Simplex Method with the Additional Rule implanted, will give us an optimal solution in at most $\binom{3n}{n}$ steps.

Chapter V

"The Method of Frank and Wolfe".

(5.1) Introduction

The method of Frank and Wolfe is a computational technique developed by Marguerite Frank and Philip Wolfe, for solving a quadratic program in Slack Form II:

$$\left.\begin{array}{ll} \text{maximise} \quad f(X) = CX + X'DX & (5.1) \\ \text{subject to} & \\ \left.\begin{array}{r} AX \leq b \\ X \geq 0 \end{array}\right] & (5.2) \\ \text{given that the quadratic form } X'DX \text{ is negative semi-definite} & \end{array}\right] \text{——}(S^2_{\cdot}II)$$

The method is based on the following facts (to be proved in section (5.2)): <u>If Problem ($S^2_{\cdot}II$) has an optimal solution</u>, then the problem

$$\left.\begin{array}{ll} \text{maximise} \quad z = -(V'X+\lambda'Y) & (5.3) \\ \text{subject to} & \\ \left.\begin{array}{l} AX + Y = b \\ ZDX - A'\lambda + V = -C' \\ X \geq 0,\ \lambda \geq 0,\ Y \geq 0,\ V \geq 0 \end{array}\right] & (5.4) \end{array}\right] \text{——}(A)$$

has an optimal solution and:

(a) The maximum of $z = -(V'X+\lambda'Y)$ subject to the constraints (5.4) has the value zero.

(b) For any optimal solution $[X_o,\lambda_o,Y_o,V_o]$, say, of Problem (A), X_o is an optimal solution to Problem ($S^2_{\cdot}II$).

(c) At least one of the basic feasible solutions of Problem (A) is an optimal solution to Problem (A).

Once the above facts are known we see that the problem of solving the quadratic program $(S^2_{\cdot}II)$ can be reduced to the following: Among the basic feasible solutions of Problem (A) find one that causes $z = -(V'X+\lambda'Y)$ to vanish. The Method of Frank and Wolfe does exactly this. It should be noted that, up to this point, the formulation is much the same as that of Wolfe's Method described in Chapter IV. Wolfe, however, proceeds by introducing artificial variables in the system (4.1), which corresponds to (5.4), and then reducing the sum of the artificial variables to zero over a sequence of basic feasible solutions, by using Phase I of the Two-Phase Simplex Method in such a way that the condition $V'X = 0$ which corresponds to (5.5.c), is satisfied at each iteration. On the other hand, the method of Frank and Wolfe proceeds in a somewhat opposite direction:- Starting with a basic feasible solution to (5.4) which does not satisfy (5.5.c) (if it did satisfy (5.5.c) we would immediately have an optimal solution to $(S^2_{\cdot}II)$), reduce the quantity $z = -(V'X+\lambda'Y)$ to zero over a sequence of basic feasible solutions without attempting to maintain $V'X + \lambda'Y = 0$, by using the gradient interpolation technique. This technique is an iterative procedure in which the principal computation is the simplex-method change of basis, and proceeds in the following way (Very brief description; the full description to be given in section (5.3)):- In the initial iteration step it requires an initial feasible solution (point) of Problem (A) (not necessarily basic) W_1, say. Once given W_1 it then selects, by the simplex routine, a secondary basic feasible solution (point) of Problem (A) Z_1, say, whose projection along the gradient of the objective function $z = -(V'X+\lambda'Y)$ at the initial point W_1, is sufficiently large. The point, which maximizes z over the line segment joining W_1 and Z_1, is then chosen as the initial point for the next iteration step, and so on ... . The values of the objective function on the initial points, thus obtained, will be shown to converge to zero. But a remarkable feature of the Frank and Wolfe Method is that this convergence is finite, i.e., in some iteration step of the gradient-interpolation algorithm, a secondary point which is an optimal solution to Problem (A) will be found, ensuring the termination of the process.

Notation. In the rest of this chapter let:

(a) $\mathbb{K}_S$ denote the feasible domain of Problem $(S^2_\cdot II)$, i.e., let $\mathbb{K}_S = \{X \mid AX \leq b,\ X \geq 0\}$.

(b) $\mathbb{K}_A$ denote the feasible domain of Problem (A), i.e., let $\mathbb{K}_A$ denote the set of all $[X,\lambda,Y,V]$ which satisfy (5.4).

(5.2) Theory of the Method.

In Chapter III Section (3.6) we obtained a very important result (see (3.13)) which, in this section, we shall refer to as Result I:

Result I. X is an optimal solution to Problem $(S^2_\cdot II)$ if and only if for some λ, Y and V, $[X,\lambda,Y,V]$ satisfies the system

$$\left.\begin{aligned} &AX + Y = b && (5.5.a)\\ &2DX - A'\lambda + V = -C' && (5.5.b)\\ &V'X + \lambda'Y = 0 && (5.5.c)\\ &X \geq 0,\ \lambda \geq 0,\ Y \geq 0,\ V \geq 0 && (5.5.d) \end{aligned}\right] \text{———} (5.5)$$

THM. I Problem $(S^2_\cdot II)$ has an optimal solution if and only if Problem (A) has a feasible solution.

PROOF: (See next page).

PROOF: Necessity Let X_o be an optimal solution to Problem ($5\overset{2}{.}II$). Then by Result I, there exists λ_o, Y_o and V_o, say, such that $[X_o, \lambda_o Y_o, V_o]$ satisfies (5.5.a), (5.5.b) and (5.5.d). Hence $[X_o, \lambda_o, Y_o, V_o]$ satisfies (5.4) and so is a feasible solution to Problem (A).

Sufficiency. Let $Z_* = [X_* \lambda_* Y_* V_*]$ be a feasible solution to Problem (A). Then,

$$\left.\begin{aligned} & AX_* + Y_* = b && (5.6.a) \\ & 2DX_* - A'\lambda_* + V_* = -C' && (5.6.b) \\ & X_* \geq 0,\ \lambda_* \geq 0,\ Y_* \geq 0,\ V_* \geq 0 && (5.6.c) \end{aligned}\right] \text{———} (5.6)$$

Now by ($THM.I\overset{*}{.}1$) we have that

$$f(X) - f(X_*) \leq \nabla f(X_*)(X - X_*)$$

for all $X \in \mathbb{K}_S = \{X | AX \leq b,\ X \geq 0\}$, since $f(X)$ is concave in E_n. Therefore

$$\begin{aligned} f(X) - f(X_*) &\leq (C + 2X_*'D)(X - X_*) && \text{since } \nabla f(X_*) = C + X_*'D \\ &= (\lambda_*'A - V_*')(X - X_*) && \text{by } (5.6.b) \\ &= \lambda_*'(AX - AX_*) - V_*'X + V_*'X_* \\ &= \lambda_*'(AX - b + Y_*) - V_*'X + V_*'X_* && \text{by } (5.6.a) \\ &\leq \lambda_*'Y_* - V_*'X + V_*'X_* && \text{since } AX \leq b \text{ and } \lambda_* \geq 0 \\ &\leq \lambda_*'Y_* + V_*'X_* && \text{since } V_* \geq 0, \text{ and } X \geq 0 \end{aligned}$$

Thus $f(X) \leq f(X_*) + \lambda_*'Y_* + V_*'X_*$, a constant, for all $X \in \mathbb{K}_S$, and hence $f(X)$ is bounded above on the feasible domain $\mathbb{K}_S$. Therefore, by THM.(VI.3) the quadratic function $f(X) = CX + X'DX$ attains its supremum (maximum) in the feasible domain $\mathbb{K}_S$, i.e., ($S\overset{2}{.}II$) has an optimal solution.

Q.E.D.

<u>THM. II</u> If Problem ($S^2_{\cdot}II$) has an optimal solution then Problem (A) has an optimal solution and

$$\underset{[X,\lambda,Y,V] \in \mathbb{K}_A}{\text{maximum}} \; z = (V'X+\lambda'Y) = 0$$

PROOF: <u>Necessity</u>. Let X_o be an optimal solution to Problem ($S^2_{\cdot}II$). Then by Result I there exists λ_o, Y_o and V_o, say, such that $[X_o,\lambda_o,Y_oV_o]$ satisfies (5.5.a), (5.5.b), (5.5.c) and (5.5.d). From (5.5.c) we have that

$$z_o = -(V_o'X_o + \lambda_o'Y_o) = 0 \; , \tag{5.7}$$

and from (5.5.a), (5.5.b) and (5.5.d) we conclude that $[X_o,\lambda_o,Y_o,V_o]$ is a feasible solution to Problem (A), i.e., $[X_o,\lambda_o,Y_o,V_o] \in \mathbb{K}_A$. But

$$z = -(V'X+\lambda'Y) \leq 0 = z_o$$

for all $[X,\lambda,Y,V] \in \mathbb{K}_A$, since $[X,\lambda,Y,V] \geq 0$ for any $[X,\lambda,Y,V] \in \mathbb{K}_A$. Hence $[X_o,\lambda_o,Y_o,V_o]$ is an optimal solution to Problem (A) and by (5.7) the value of the objective function $z = -(V'X+\lambda'Y)$ at $[X_o,\lambda_o,Y_o,V_o]$ ($\in \mathbb{K}_A$) is zero.

Q.E.D.

<u>THM.III.</u> If Problem (A) has an optimal solution then Problem ($S^2_{\cdot}II$) has an optimal solution and

$$\underset{[X,\lambda,Y,V] \in \mathbb{K}_A}{\text{maximum}} \; z = -(V'X+\lambda'Y) = 0 \tag{5.8}$$

PROOF: Let $[X_o,\lambda_oY_o,\lambda_o]$ be an optimal solution to Problem (A). Then, clearly, $[X_o,\lambda_o,Y_o,V_o]$ is a feasible solution to Problem (A) and hence, by THM. I, Problem ($S^2_{\cdot}II$) has an optimal solution which implies (5.8) by THM. II.

Q.E.D.

THM. IV. Problem (S^2.II) has an optimal solution if and only if Problem (A) has an optimal solution.

PROOF: Follows immediately from THM. II and III.

THM. V. If Problem (S^2.II) has an optimal solution then:

(i) Problem (A) has an optimal solution.

(ii) $\underset{[X,\lambda,Y,V] \in \mathbb{K}_A}{\text{Maximum}} \; z = -(V'X+\lambda'Y) = 0$

(iii) For any optimal solution $[X_o,\lambda_o,Y_o,V_o]$ of Problem (A), X_o is an optimal solution to Problem (S^2.II).

PROOF: (i) and (ii) follow immediately from THM. II. For the proof of (iii) let $[X_o,\lambda_o,Y_o,V_o]$ be an optimal solution to Problem (A). Then by (ii)

$$Y_o'X_o + \lambda_o'Y_o = 0$$

Also

$$AX_o + Y_o = b$$

$$2DX_o - A'\lambda_o + V_o = -C'$$

$$X_o \geq 0,\ \lambda_o \geq 0,\ Y_o \geq 0,\ V_o \geq 0\ ,$$

since $[X_o,\lambda_o,Y_oV_o]$ is a feasible solution to Problem (A). Hence (iii) follows from Result I.

Q.E.D.

THM. VI. If Problem (A) has a feasible solution then:

(i) Problem (A) has an optimal solution.

(ii) $\underset{[X,\lambda,Y,V] \in \mathbb{K}_A}{\text{Maximum}} \; z = -(V'X+\lambda'Y) = 0$

(iii) For any optimal solution $[X_o,\lambda_o,Y_o,V_o]$ of Problem (A), X_o is an optimal solution to Problem (S^2.II).

PROOF: (See next page).

PROOF: Since Problem (A) has a feasible solution, Problem (S^2.II) has an optimal solution by THM. I. Hence results follows immediately by THM. V.

Q.E.D.

REMARK: This theorem shows that Problem (A) does not have unbounded solutions, i.e., Problem (A) either has no feasible solution or then it has an optimal solution.

Now Problem (A) either has a feasible solution or has no feasible solution:

(a) If Problem (A) has a feasible solution then:

(i) Problem (A) has an optimal solution (THM. VI.).

(ii) Problem (S^2.II) has an optimal solution (THM. I).

(b) If Problem (A) has no feasible solution then Problem (S^2.II) cannot have an optimal solution. For if Problem (S^2.II) had an optimal solution, Problem (A) would have a feasible solution (THM. I) contradicting our assumption. But from section (3.10) a quadratic programming problem can only have either:

(i) An optimal feasible solution, or

(ii) An unbounded solution, or

(iii) No feasible solution.

Hence, we conclude that, <u>if Problem (A) has no feasible solution then either Problem (S^2.II) has no feasible solution or then it has an unbounded solution.</u>

We are now in position to describe the Frank and Wolfe Algorithm, but before we do so it is convenient to write Problem (A) in a more compact form:-

If we combine the four interdependent vectors X,λ,Y,V into a single vector Z, then the constraints (5.4) of Problem (A) can be written in the form

$$BZ = d$$

$$Z \geq 0$$

where (i) B is an $(m+n)\times 2(m+n)$ matrix, namely

$$B = \begin{bmatrix} A & 0 & I_m & 0 \\ 2C & -A' & 0 & I_n \end{bmatrix}$$

(I_m denotes the identity matrix of rank m and I_n the identity matrix of rank n).

(ii) Z is a $2(m+n)$ column vector, namely $Z = [X,\lambda,Y,V]$

(iii) d is an $(m+n)$ column vector, namely $d = [b,-C']$.

If we next let $\overline{Z}$ denote the "adjoint" of Z, i.e., if we put

$$\overline{Z} = [V';Y';\lambda';X'] \quad ,$$

then the objective function $z = -(V'X+\lambda'Y)$ of Problem (A) becomes

$$z = -\frac{1}{2}\overline{Z}Z$$

Hence with the above notation we can write Problem (A) in the form

$$\left.\begin{array}{ll} \text{maximize } T(Z) = -\overline{Z}Z & (5.9) \\ \text{subject to} & \\ \left.\begin{array}{l} BZ = d \text{ ———— } (5.10.a) \\ Z \geq 0 \text{ ———— } (5.10.b) \end{array}\right] & (5.10) \end{array}\right] \text{———— } (A^*)$$

REMARKS:

(i) Note that in (5.9) we have dropped the factor $\frac{1}{2}$. This in no way affects our set up and spares us the trouble of keeping track of it in each iteration step of the algorithm.

(ii) For any $Z_* = [X_*,\lambda_*,Y_*,V_*]$ and $Z = [X,\lambda,Y,V]$ satisfying (5.10.a) we have that:

(a) $\overline{Z}_*Z = \overline{Z}Z_*$ (5.11)

PROOF: Trivial

(b) $-\frac{1}{2}(\overline{Z}_*-\overline{Z})(Z_*-Z) = 2(X_*-X)'D(X_*-X) \leq 0$ (5.12)

PROOF:

$$\frac{1}{2}(\overline{Z}_*-\overline{Z})(Z_*-Z) = (V_*'-V')(X_*-X) + (\lambda_*'-\lambda')(Y_*-Y) \qquad \text{using (5.11)}$$

$$= \left[-2(X_*-X)'D+(\lambda_*-\lambda)'A\right](X_*-X)-(\lambda_*-\lambda)'A(X_*-X) \qquad \text{by (5.10.a)}$$

$$= -2(X_*-X)'D(X_*-X)$$

Hence (5.12) follows since $X'DX$ is a negative semi-definite quadratic form.

(iii) (5.12) reflects the concativity of the objective function $T(Z)$ over the feasible domain of Problem (A^*).

(iv) Note that we still haven't proved that, when Problem $(S^2_\cdot II)$ has an optimal solution (or equivalently when Problem (A^*) has a feasible solution) at least one of the basic feasible solutions of Problem (A^*) will be an optimal solution to Problem (A^*). This we will do in Section (5.4) when we prove the finiteness of the Frank and Wolfe Algorithm.

(5.3) The Frank and Wolfe Algorithm.

Although this has not been necessary before, in this section, for the application of the simplex method, we suppose that the constraint equations (5.10.a) have, were necessary, been multiplied by -1, so that the right hand side is positive, i.e., so that $d \geq 0$.

Phase I: In this phase the constraints (5.10.a) are tested for feasibility by applying Phase I of the Two-Phase Simplex Method (see App.V.1). If the constraints (5.10.a) are feasible, a basic feasible solution $Z_\dagger$ is produced with which to begin Phase II. If the constraints (5.10.a) are not feasible then the last n equations of system (5.10.a) may be discarded and the remaining system of constraints

$$AX + Y = b \ .$$

is similarly tested for feasibility. If these constraints are feasible, then the quadratic program $(S^2_{\cdot}II)$ is feasible but unbounded above. Otherwise the quadratic problem $(S^2_{\cdot}II)$ is infeasible (inconsistent).

Phase II: This Phase is defined inductively. Suppose that in the kth iteration step we have:

(i) A basic feasible solution to Problem (A^*) say, Z_K, with an associated simplex tableau so as to be able to start the simplex machinery.

(ii) A feasible solution to Problem (A^*) (not necessarily basic) say, W_K, for which $T(W_K) = -\overline{W}_K W_K < 0$. (If $T(W_K) = 0$ then W_K would be optimal and we need not proceed any further.

[Note. To start Phase II use the basic feasible Z_+ obtained at the end of Phase I as Z_1 and W_1, i.e., in the first iteration step of Phase II let $Z_1 = Z_+$ and $W_1 = Z_+$.]

With Z_K as an initial basic feasible solution, use the simplex algorithm to maximise the linear form

$$L_K(Z) = -\overline{W}_K Z$$

subject to the constraints

$$BZ = d$$

$$Z \geq 0 \ .$$

We then obtain a sequence of basic feasible solutions to Problem (A^*)

$$Z_K^1 = Z_K \ , \quad Z_K^2 \ , \quad Z_K^3 \quad , \ \ldots$$

with $\quad -\overline{W}_K Z_K^1 < -\overline{W}_K Z_K^2 < -\overline{W}_K Z_K^3 < \ldots$ (see App.V.2)

Stop the simplex procedure at the first Z_K^h for which either

$$\text{(a) } \overline{Z}_K^h Z_K^h = 0 \tag{5.13}$$

$$\text{or} \quad \text{(b) } -\overline{W}_K Z_K^h \geq -\frac{1}{2}\overline{W}_K W_K , \tag{5.14}$$

is satisfied.

If (5.13) is obtained, Z_K^h is an optimal solution to Problem (A).

If (5.14) is obtained let

$$\text{(i) } Z_{K+1} = Z_K^h$$

$$\text{(ii) } W_{K+1} = W_K + \mu(Z_K^h - W_K)$$

where $$\mu = \text{Min } \left\{\frac{\overline{W}_K(W_K - Z_K^h)}{(\overline{Z}_K^h - \overline{W}_K)(Z_K^h - W_K)}, 1\right\} \tag{5.15}$$

Repeat Phase II, using W_{K+1} and Z_{K+1} .

REMARK: Note that the Frank and Wolfe Algorithm, as the Simplex Method, requires an initial basic feasible solution to get started.

(5.4) Convergence of the Frank and Wolfe Algorithm.

In this section we shall prove:

(i) That in each iteration step of Phase II either (5.13) or (5.14) must occur; and

(ii) That (5.13) must occur once after finitely many steps, so that a solution will have been found. This is equivalent to showing that in some iteration step of Phase II a basic optimal feasible solution to Problem (A^*) will be obtained (note that this is the only fact mentioned in section (5.1) that we still have not proved).

PROOF OF (i):

To prove that a Z_K^h satisfying (5.14) will be found in the Kth ($K = 1,2,\ldots$) iteration step, if (5.13) is not obtained , note that the linear programming problem

$$\left.\begin{array}{l} \text{maximize } L_K(Z) = -\overline{W}_K Z \\ \text{subject to} \\ \left.\begin{array}{r} BZ \leq d \\ Z \geq 0 \end{array}\right] \qquad (5.16) \end{array}\right] \text{——} (L_K^*)$$

has a finite maximum (i.e. it is not unbounded), since $-\overline{W}_K Z \leq 0$ (see App.V.3), and that, since we are in Phase II, Problem (A^*) has a feasible solution and hence an optimal solution Z_*, say (THM. VI). So, since for any optimal solution of Problem (A^*) the objective function vanishes (THM. VI) we conclude that $T(Z_*) = -\overline{Z}_* Z_* = 0$. Therefore

$$\begin{aligned} \overline{W}_K W_K - 2\overline{W}_K Z_* &= \overline{W}_K W_K - 2\overline{W}_K Z_* + \overline{Z}_* Z_* \\ &= (\overline{W}_K - \overline{Z}_*)(W_K - Z_*) \\ &\geq 0 \qquad\qquad \text{by } (5.12) \end{aligned}$$

That is

$$-\overline{W}_K Z_* \geq -\frac{1}{2}\overline{W}_K W_K .$$

But Z_* is a feasible solution to Problem (L_K^*) and hence we must have that the finite maximum of this problem is $\geq -\frac{1}{2}\overline{W}_K W_K$. Whence, since the simplex algorithm reaches the maximum of any linear programming problem in a finite number of iterations we have our result.

PROOF OF (ii):

To prove the finite convergence of the Frank and Wolf algorithm we first prove that the function $-\overline{W}_K W_K$ converges monotonically to zero with each iteration step of the algorithm and then use this fact to finally prove the finite convergence of the method.

(a) Proof that $-\overline{W}_K W_K \to 0$ as $K \to \infty$.

We begin by noting that for μ, as defined in (5.15), we have $0 < \mu \leq 1$. This follows because

Firstly:

$$\overline{W}_K(W_K - Z_K^h) \geq \frac{1}{2}\overline{W}_K W_K \quad \text{(since by (5.14) } -\overline{W}_K Z_K^h \geq -\frac{1}{2}\overline{W}_K W_K)$$

$$> 0 \,, \qquad (5.17)$$

for $\overline{W}_K W_K = 0$ would imply $W_K = 0$ and since W_K is a feasible solution to problem (A^*) this would mean that the "requirements vector" $d = [b, -C']$ of Problem (A^*) would be equal to the zero vector, contradicting our stipulation made in section (3.1) that $b \neq 0$.

And Secondly:

$$(\overline{Z}_K^h - \overline{W}_K)(Z_K^h - W_K) \geq 0 \qquad \text{by (5.12)}$$

$$> 0 \qquad \text{since otherwise we would have}$$

$Z_K^h - W_K = 0$, contradicting (5.17).

Hence $0 < \mu \leq 1$ and so $W_{K+1} = W_K + \mu(Z_K^h - W_K)$ is a convex combination of W_K and Z_K^h (indeed it has been chosen to as to maximise the objective function $T(Z)$ on the line segment joining W_K and $Z_K^h (= Z_{K+1})$).

Next we prove that when Z_K^h satisying $-\overline{W}_K Z_K^h \geq -\frac{1}{2}\overline{W}_K W_K$ is obtained in the Kth step of Phase II we then have, for the $(K+1)$st step of Phase II, that:

$$\begin{aligned}
\overline{W}_{K+1}W_{k+1} &= \overline{W}_K W_K + 2\mu\overline{W}_K(Z_K^h - W_K) + \mu^2(\overline{Z}_K^h - \overline{W}_K)(Z_K^h - W_K) \\
&= \overline{W}_K W_K + \mu\overline{W}_K(Z_K^h - W_K) + \mu\left[\mu(\overline{Z}_K^h - \overline{W}_K)(Z_K^h - W_K) - \overline{W}_K(W_K - Z_K^h)\right] \\
&\leq \overline{W}_K W_K + \mu\overline{W}_K(Z_K^h - W_K) \qquad \text{by } (5.15) \\
&\leq \overline{W}_K W_K - \mu\frac{1}{2}\overline{W}_K W_K \qquad \text{since } -\overline{W}_K Z_K^h \geq -\frac{1}{2}\overline{W}_K W_K \\
&= (1 - \tfrac{1}{2}\mu)\overline{W}_K W_K \quad . \qquad (5.18)
\end{aligned}$$

Next let $\mathbb{B}$ be the set of all basic feasible solutions to Problem (A^*) and let $\mathbb{Y}$ be the set of all convex combinations of the points of $\mathbb{B}$. Since the number of basic solutions is always finite, we may suppose that $\mathbb{B}$ has N elements, say. Hence $\mathbb{Y}$ can be written as

$$\mathbb{Y} = \{\sum_{i=1}^{N} t_i Z_i \,|\, Z_i \in \mathbb{B} \text{ and } t_i \geq 0 \text{ for } i = 1,\ldots,N; \; \sum_{i=1}^{N} t_i = 1\}$$

Therefore $\mathbb{Y}$ is a convex polyhedron spanned by the basic feasible solutions of Problem (A^*) and hence is closed and bounded (see App. V.4).

Now, if we let

$$L = \text{Max}\,\{(\overline{Z}_* - \overline{Z}_+)(Z_* - Z_+) \,|\, Z_*, Z_+ \text{ in } \mathbb{B}\}$$

then $L > 0$ by (5.12) and

$$\frac{1}{2}\mu \underset{(1)}{\geq} \frac{\overline{W}_K(W_K - Z_K^h)}{2L} \underset{(2)}{\geq} \frac{\overline{W}_K WK}{4L}, \quad \underline{\text{if } \mu << 1}$$

((1) follows from (5.15) and (2) from (5.14)). Hence

$$1 - \frac{\mu}{2} \leq \text{Max}\,\{1 - \frac{\overline{W}_K W_K}{4L}, \frac{1}{2}\} \quad \text{for } 0 < \mu \leq 1 \quad ,$$

since if $\mu = 1$ then $\frac{\mu}{2} = \frac{1}{2}$. And so by (5.18)

$$\frac{\overline{W}_{K+1}W_{K+1}}{4L} \leq \frac{\overline{W}_K W_K}{4L} \text{ Max}\,\{1 - \frac{\overline{W}_K W_K}{4L}, \frac{1}{2}\} \qquad (5.19)$$

If we put $a_K = \frac{\overline{W}_K W_K}{4L}$, then relation (5.19) can be written in the more compact form

$$a_{K+1} \leq \text{Max}\,\{1-a_K, \tfrac{1}{2}\}a_K \qquad \text{for } K \geq 1 \tag{5.20}$$

Now

(i) If $a_K \geq \frac{1}{2}$ for all $K \geq 1$ then $a_{K+1} \leq \frac{1}{2} a_K$ by (5.20) and therefore

$$a_{K+1} \leq (\tfrac{1}{2})^K a_1$$

Hence $a_K \to 0$ as $K \to \infty$ i.e. $-\overline{W}_K W_K$ 0 as $K \to \infty$.

(ii) If $a_K < \frac{1}{2}$ for some value of K say, K_*, i.e. $a_{K_*} < \frac{1}{2}$ then $a_{K_*+1} \leq (1-a_{K_*})a_{K_*}$ by (5.20) and therefore $a_{K_*+1} \leq a_{K_*} - a_{K_*}^2 < \frac{1}{2}$.

Hence if $a_K < \frac{1}{2}$ for some value of K, say K_*, then $a_K < \frac{1}{2}$ for all $K \geq K_*$. But from (5.20) we have that

$$a_{K+1} \leq (1-a_K)a_K \quad \text{whenever } a_K < \tfrac{1}{2}.$$

<u>Whence if $a_K < \frac{1}{2}$ for some value of K, say K_*, then</u>

$$a_{K+1} \leq (1-a_K)a_K \quad \text{for all } K \geq K_*$$

i.e. $$\frac{1}{a_{k+1}} \geq \frac{1}{a_K} \cdot \frac{1}{1-a_K} = \frac{1+a_K+a_K^2+\ldots}{a_K} \geq \frac{1}{a_K} + 1 \quad \text{for all } K \geq K_*$$

so that

$$\frac{1}{a_K} \geq \frac{1}{a_{K_*}} + (K - K_*) \qquad \text{for all } K \geq K_*.$$

Thus

$$a_K \leq \frac{1}{(K-K_*) + \frac{1}{a_{K_*}}} \qquad \text{for all } K \geq K_*.$$

Hence $a_K \to 0$ as $K \to \infty$, i.e., $-\overline{W}_K W_K \to 0$ as $K \to 0$

Hence from (i) and (ii) above we conclude that $\overline{W}_K W_K \to 0$ as $K \to \infty$ and the convergence of the Frank and Wolfe Algorithm is proved.

(b) <u>Proof of the finite convergence of the Frank and Wolf Algorithm.</u>

Suppose that there does not exist a value (finite) of K for which the linear programming problem (L_K) of Phase II yields a basic feasible solution which is an optimal solution to Problem (A^*), i.e., suppose that, as K runs from 1 to ∞, the linear programming problem (L_K) never yields a basic feasible solution which is an optimal solution to Problem (A^*). Then, in particular, none of the Z_K which is obtained at the end of <u>each</u> iteration step of Phase II, is an optimal solution to Problem (A^*) and so $W_\infty = 0$ is a boundary point of

$$\mathbb{Y} = \{ \sum_{i=1}^{N} t_i Z_i \mid Z_i \in \mathbb{B} \text{ and } t_i \geq 0 \text{ for } i = 1,\ldots,N; \sum_{i=1}^{N} t_i = 1\}$$

This is justified in the following way:

$W_\infty = 0$ is a boundary (limit) point of the infinite set $\mathbb{W} = \{W_K \mid K=1,2,\ldots,\}$ since $\lim\limits_{K\to\infty} \overline{W}_K W_K = 0$. The set $\mathbb{W}$ is a subset of $\mathbb{Y}$ since each W_K is a convex combination of $\{Z_i \mid i < K\}$. Hence $W_\infty \in \mathbb{Y}$. (In relation to $\mathbb{Y}$, W_∞ can either be exterior, interior or boundary. If it is exterior then it cannot be a boundary point of $\mathbb{W}$ since $\mathbb{W} \subseteq \mathbb{Y}$. Therefore W_∞ is either an interior point or a boundary point of $\mathbb{Y}$. But $\mathbb{Y}$ is closed therefore $W_\infty \in \mathbb{Y}$). Now

$$\mathbb{Y} = \{ \sum_{i=1}^{N} t_i Z_i \mid Z_i \in \mathbb{B} \text{ and } t_i \geq 0 \text{ for } i = 1,\ldots,N; \sum_{i=1}^{N} t_i = 1\}$$

and therefore we have that

$$\begin{aligned}\overline{W}_\infty W_\infty &= \left(\sum_{i=1}^{N} t_i \overline{Z}_i\right)\left(\sum_{i=1}^{N} t_i Z_i\right)\\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} t_i t_j \overline{Z}_i Z_j\\ &> 0\end{aligned}$$

Multiply the constraints of the system $BZ = d$ by -1 (where necessary) so that $d \geq 0$. Initiate Phase I.

(b) Phase I.

Step 1. Test the system $BZ = d$ $(d \geq 0)$ for feasibility using Phase I of the Two-Phase Simplex Method:

(i) If a basic feasible solution $Z_{\dagger}$, say, to $BZ = d$ is obtained, i.e., if the system $BZ = d$ is feasible, initiate Phase II with $W_K = Z_{\dagger}$ and $Z_K = Z_{\dagger}$, for Problem $S^2.II$ has an optimal solution.

(ii) If a basic feasible solution to $BZ = d$ is not obtained i.e., if the system $BZ = d$ is not feasible pass to Phase I Step 2, for Problem $S^2.II$ has no optimal solution.

Step 2. Test the system $AX + Y = b$ $(b \geq 0)$ for feasibility using Phase I of the Two-Phase Simplex Method:

(i) If a basic feasible solution to $AX + Y = b$ is found then Problem $S^2.II$ has an unbounded solution.

(ii) If a basic feasible solution to $AX + Y = b$ is not obtained then Problem $S^2.II$ has no feasible solution.

(c) Phase II. Given a feasible solution W_K, say, and a basic feasible solution Z_K, say, to the system $BZ = d$, obtain a sequence of "improving" basic feasible solutions $Z_K^1 = Z_K, Z_K^2, Z_K^3, \ldots, Z_K^h, \ldots$ to the linear programming problem

$$\text{maximise} \quad L_K(Z) = -\overline{W}_K Z$$

subject to

$$BZ = d$$
$$Z \geq 0 ,$$

by the use of the simplex algorithm. Stop at the first Z_K^h for which either

$$\text{(i)}\quad \overline{Z}_K^h Z_K^h = 0$$

$$\text{or}\quad \text{(ii)}\quad -\overline{W}_K Z_K^h \geq -\frac{1}{2}\overline{W}_K W_K$$

<u>If (i) is obtained</u> then the X-component of $Z_K^h = [X_K^h, Y_K^h, \lambda_K^h, V_K^h]$, namely, X_K^h is an optimal solution to Problem (S^2.II).

<u>If (ii) is obtained</u> let

(a) $Z_{K+1} = Z_K^h$

(b) $W_{K+1} = W_K + \mu(Z_K^h - W_K)$ where

$$\mu = \text{Min}\left\{\frac{\overline{W}_K(W_K - Z_K^h)}{(\overline{Z}_K^h - \overline{W}_K)(Z_K^h - W_K)}, 1\right\},$$

and repeat Phase II with W_K and Z_K replaced by W_{K+1} and Z_{K+1} respectively.

<u>Computation Time</u>: Experience with this method suggests that the total number of changes of basis (several of which occur in each iteration step of Phase II) required to solve the quadratic program in Slack Form II will, in practice, be of the order of $2(m+n)$. Each iteration step of Phase II after the first, commonly require very few changes of basis (in numerical examples of the order $n = 5$, $m = 3$ we usually have only one or two changes of basis in each iteration step of Phase II). Thus the Frank and Wolfe Algorithm seems to solve quadratic problems as quickly as the Simplex Method solves linear problems of comparable size.

We end this chapter with two important remarks about the Frank and Wolfe Method:

Remarks: (i) Note that a computer program for the Simplex Method requires only very few alterations to be adapted for the Frank and Wolfe Method.

(ii) Note also that

$$\frac{\partial T}{\partial Z} = -2\overline{Z}$$

and, in particular

$$\left[\frac{\partial T}{\partial Z}\right]_{Z = W_K} = -2\overline{W}_K$$

Therefore the linear function $L_K(Z) = -\overline{W}_K Z$ is obtained by linearizing $T(Z) = -\overline{Z}Z$ at the point W_K, so that the gradient of these two functions at the point W_K have the same direction. This should help in making the Method of Frank and Wolfe "intuitively" more clear.

Chapter VI

"Dantzig's Method"

(6.1) Introduction.

Dantzig's Method is a computational technique developed by George B. Dantzig, for solving a quadratic program in Normal Form II:

$$\left.\begin{array}{l} \text{maximise} \quad f(X) = CX + X'DX \\ \text{subject to} \\ \qquad AX = b \\ \qquad X \geq 0\ , \\ \text{given that the quadratic form } X'DX \text{ is negative semi-definite.} \end{array}\right] \qquad (N^2_{\circ}II)$$

The method is based on the following facts (to be proved in section (6.2)):

(i) There exists an X,λ,V satisfying the system of equations

$$\begin{bmatrix} A & 0 & 0 \\ 2D & -A' & I_n \end{bmatrix} \begin{bmatrix} X \\ \lambda \\ V \end{bmatrix} = \begin{bmatrix} b \\ -C' \end{bmatrix} \qquad (6.1)$$

and the additional conditions

$$\left.\begin{array}{l} \left.\begin{array}{l} X \geq 0 \\ V \geq 0 \\ \lambda \ \ \text{unrestricted} \end{array}\right] \text{———} (6.2.a) \\ V'X = 0, \text{———} (6.2.b) \end{array}\right] \qquad (6.2)$$

if and only if Problem $(N^2_{\circ}II)$ has an optimal solution.

(ii) For any $[X,\lambda,Y]$ satisfying (6.1) and (6.2), X is an optimal solution to Problem $(N^2_{\circ}II)$.

(iii) If Problem $(N^2_{\circ}II)$ has an optimal solution there exists at least one basic solution $[X,\lambda,Y]$ to (6.1) satisfying (6.2).

If you recall, this is exactly the same "set up" as the one we had in section (4.1) when we introduced Wolfe's Method. Dantzig's Method is indeed a variant of Wolfe's elegant procedure. The chief difference is that Dantzig's Method is "more nearly" a strict analogue of the Simplex Method; it has a "tighter" selection rule and a monotonically increasing objective. These differences will become more clear once we have presented the method.

(6.2) Theory of the Method.

In Chapter III section (3.6) we proved the following result (see relation (3.9)):

Result I. X is an optimal solution to Problem ($N^2_{\cdot}II$) if and only if for some λ and V, $[X,\lambda,V]$ satisfies

$$\begin{aligned} &AX = b \\ &2DX - A'\lambda + V = -C' \\ &V'X = 0 \\ &X \geq 0,\quad V \geq 0,\ \lambda \text{ unrestricted} \end{aligned}$$

(this proves statements (i) and (ii) made in the previous section). Hence if Problem ($N^2_{\cdot}II$) has an optimal solution then there exists an $[X,\lambda,V]$ satisfying (6.1), (6.2.a) and (6.2.b). But by THM. (X.3) provided Problem ($N^2_{\cdot}II$) has an optimal solution there exists at least one $|X,\lambda,Y|$ satisfying (6.1) and (6.2.b) which is a basic solution to system (6.1) (this proves statement (iii) made in the previous section), whence:

Result II. If Problem ($N^2_{\cdot}II$) has an optimal solution then there exists a basic solution to system (6.1), say, $[X,\lambda,V]$, satisfying (6.2.a) and (6.2.b), whose X-component is an optimal solution to Problem ($N^2_{\cdot}II$).

Now consider the following system

$$\left.\begin{array}{l} \left.\begin{array}{l} AX = b \\ 2DX - A'\lambda + V = -C' \end{array}\right] \text{———} \ 6.3.a \\ X \geq 0 \ \text{———} \ 6.3.b \\ \lambda \ \text{unrestricted} \ \text{———} \ 6.3.c \\ V \ \text{unrestricted} \ \text{———} \ 6.3.d \end{array}\right] \qquad (6.3)$$

where A, D, b, $-C'$ are as defined for system (6.1) and where as before:

$$X = [x_1, \ldots\ldots, x_n]$$

$$\lambda = [\lambda_1, \ldots, \lambda_m]$$

$$V = [v_1, \ldots\ldots, v_n]$$

In relation to system (6.3) make the following definitions:

DEFN. Feasible Solution. Any solution $[X,\lambda,V]$ to (6.3.a) which also satisfies (6.3.b) is called a feasible solution to system (6.3) (note that neither λ or V are sign restricted for any feasible solution $[X,\lambda,V]$ to (6.3)).

DEFN. Basic Feasible Solution. Any basic solution $[X,\lambda,V]$ to (6.3.a) which also satisfies (6.3.b) is called a basic feasible solution to system (6.3).

DEFN. Complementary Basic Feasible Solution. A basic feasible solution $[X,\lambda,V]$ to system (6.3) is said to be complementary if, for each j either x_j or v_j, but not both, is a basic variable, i.e., if, for each j either x_j or v_j, but not both, is in the basic set.

Note. In what follows the system (6.3) will be assumed to be nondegenerate. This in no way will imply any loss of generality (see Appendix VI.1).

Now by Result I above, the vector (point) X_o constitutes an optimal solution to Problem (N^2.II) if there exists λ_o and V_o such that:

(1) $[X_o,\lambda_o,V_o]$ satisfies (6.3.a)

(2) X_o satisfies (6.3.b)

(3) λ_o satisfies (6.3.c)

(4) V_o is <u>non-negative</u>, i.e., $V_o \geq 0$

(5) $V_o'X_o = 0$ (i.e., $v_j^o = 0$ if $x_j^o > 0$) .

Hence we see that the X-component of a feasible solution $[X,\lambda,V]$ to system (6.3) will be an optimal solution to Problem (N^2.II) if $V \geq 0$ and $V'X = 0$. But by Result II we know that, provided Problem (N^2.II) has an optimal solution, there exists a basic feasible solution $[X_o,\lambda_o,V_o]$, say, to system (6.3) satisfying $V_o \geq 0$ and $V_o'X_o = 0$. Therefore all we need, to find an optimal solution to Problem (N^2.II), is to move from basic feasible solution to basic feasible solution (of system (6.3)) until we arrive at one, say, $[X_o,\lambda_o,V_o]$, satisfying $V_o \geq 0$ and $V_o'X_o = 0$.But here, as we next proceed to show, this movement over a sequence of basic feasible solutions of system (6.3) will be made in such a way as to ensure an "increase", at each step, in the value of the objective function $f(X) = CX + X'DX$ (recall that this is one of the main characteristics of the simplex algorithm:- <u>at each iteration an "improvement" of the objective function is obtained</u>).

Suppose we select from the constant matrix of system (6.3), i.e., from

$$\begin{bmatrix} A & 0 & 0 \\ 2D & -A' & I_n \end{bmatrix} \tag{6.4}$$

a nonsingular matrix B. Denote by $Z_B = [z_{B1},\ldots,z_{B,m+n}]$ the variables of $[X,\lambda,V]$ associated with columns of (6.4) in B. Let R contain the columns of (6.4) not in B, and let $Z_R = [z_{R1},\ldots,z_{Rn}]$ denote the variables of $[X,\lambda,V]$ associated with the columns of (6.4) not in B. Then any solution to the system (6.3.a) can be written

$$Z_B = B^{-1}\begin{bmatrix} b \\ -C' \end{bmatrix} - B^{-1}RZ_R$$

A basic solution to (6.3.a) is obtained by setting $Z_R = 0$. The basic solution will be feasible if the x-components of X present in Z_B are positive. Next let

$$B^{-1}\begin{bmatrix} b \\ -C' \end{bmatrix} = Y_o = [y_{1o},\ldots,y_{no}]$$

and denote the jth column of $B^{-1}R$ by $Y_j = [y_{ij},\ldots,y_{nj}]$. Then any solution to (6.3.a) must satisfy

$$z_{Bi} = y_{io} - \sum_{j=1}^{n} y_{ij}z_{Rj} \quad i = 1,\ldots,m+n \tag{6.5}$$

The z_{Bi} $i = 1,\ldots,m+n$, are called the basic variables
The z_{Rj} $j = 1,\ldots,n$ are called the non-basic variables

Now suppose that $Z_B = Y_o$ is a basic feasible solution to the system (6.3.a) so that $Z_R = 0$. Suppose next that we start increasing in (6.5) the value of the nonbasic variable z_{Rj} for $j = s$, i.e., suppose we start increasing the nonbasic variable z_{Rs}, then at some stage the value of some basic variable, say, z_{Br}, will become zero (provided of course that $y_{is} > 0$ for at least one value of i), and a new basic solution will be obtained with z_{Rs} in the basic set i.e., with z_{Rs} as a basic variable, and with $z_{Br} = 0$, i.e., with z_{Br} as a nonbasic variable (see App. VI.2). Please note that the above procedure is exactly equivalent to the procedure of the Simplex Algorithm where the basic variable x_{Br}, which drops from the basic set upon introduction of the non-basic variable x_s into the basic set, is determined by

$$\frac{x_{Br}}{y_{rs}} = \min_{i:y_{is}>0} \{\frac{x_{Bi}}{y_{is}}\}$$

We are now in position to prove the theorems in which Dantzig's Method is based:

<u>THM. I</u> Let $[X,\lambda,V]$ and $[X_*,\lambda_*,V_*]$ be any two feasible solutions to system (6.3) then for $f(X) = CX + X'DX$ we have that

$$f(X) - f(X_*) = -V_*'(X-X_*)'D(X-X_*) \tag{6.6}$$

PROOF: We have that

$$f(X) - f(X_*) = (CX + X'DX) - (CX_*+X_*'DX_*) \tag{6.7}$$

Now from (6.3.a) we obtain

$$2DX_* - A'\lambda_* + V_* = -C' \tag{6.8}$$

Hence premultiplying relation (6.8) by X_*' and X' we get

$$2X_*'DX_* - X_*'A'\lambda_* + X_*'V_* = -X_*'C' \tag{6.9}$$

$$2X'DX_* - X'A'\lambda_* + X'V_* = -X'C' \tag{6.10}$$

Thus, since $AX_* = b$ and $AX = b$, we conclude on subtracting (6.10) from (6.9), that $2X_*'DX_* - 2X'DX_* + V_*'X_* - V_*'X = CX - CX_*$

And so since $(X-X_*)'D(X-X_*) = X'DX - 2X'DX_* + X_*'DX_*$, we finally obtain that

$$CX + X'DX -(CX_*+X_*'DX_*) = -V_*'(X-X_*) + (X-X_*)'D(X-X_*)$$

Whence result follows by (6.7)

Q.E.D.

<u>THM. II</u>. Let $[X_o,\lambda_o,V_o]$ be a complementary basic feasible solution to the system (6.3) and suppose that $v_s^o < 0$. Then any increase of the nonbasic variable x_s with adjustment of only the basic variables, generates a class of feasible solutions to system (6.3), namely,

$\mathbb{B} = \{[X,\lambda,V] = [\overline{X},\overline{\lambda},\overline{V}] \mid [\overline{X},\overline{\lambda},\overline{V}]$ a feasible solution to (6.3)$\}$,

such that $f(X) = CX + X'DX$ increases as long as $v_s = \overline{v}_s < 0$.

PROOF: Let $[X,\lambda,V]$ be any solution in the class $\mathbb{B}$, i.e. let $[\overline{X},\overline{\lambda},\overline{V}]$ be any solution generated by increasing $x_s = x_s^o = 0$ to $x_s = \overline{x}_s > 0$, say. Let $[\bar{\bar{X}},\bar{\bar{\lambda}},\bar{\bar{V}}]$ be another solution in the class $\mathbb{B}$, i.e., let $[\bar{\bar{X}},\bar{\bar{\lambda}},\bar{\bar{V}}]$ be another solution generated by increasing $x_s = x_s^o = 0$ to $x_s = \bar{\bar{x}}_s > \overline{x}_s$, say. Then, from relation (6.6),

$$f(\bar{\bar{X}}) - f(\overline{X}) = -\overline{v}_s(\bar{\bar{x}}_s-\overline{x}_s) + (\bar{\bar{X}}-\overline{X})'D(\bar{\bar{X}}-\overline{X}), \tag{6.11}$$

since for all $j \neq s$ either $\bar{\bar{x}}_j = \overline{x}_j = 0$, if x_j is a nonbasic variable or $\bar{\bar{v}}_j = \overline{v}_j = 0$, if x_j is a basic variable (by the complementary assumption).

Now from (6.5) we see that the adjusted values of the basic variables are linear functions of the nonbasic variables and so since all the non-basic variables with the exception of x_s, are being kept equal to zero we have that

$$(\bar{\bar{X}} - \overline{X}) = (\bar{\bar{x}}_s - \overline{x}_s)Y \tag{6.12}$$

where Y is a constant vector. Hence, by (6.11),

$$f(\bar{\bar{X}}) - f(\overline{X}) = (\bar{\bar{x}}_s-\overline{x}_s)\left[-\overline{v}_s + (\bar{\bar{x}}_s-\overline{x}_s)(Y'DY)\right]$$

Therefore it is clear that if $\overline{v}_s < 0$, $f(\bar{\bar{X}}) - f(\overline{X}) > 0$ for $(\bar{\bar{x}}_s-\overline{x}_s) > 0$ sufficiently near to zero, i.e., if $\overline{v}_s < 0$ then

$$f(\bar{\bar{X}}) > f(\bar{X}) \tag{6.13}$$

for $\bar{\bar{x}}_s > \bar{x}_s$ but sufficiently near to $\bar{x}_s$. But for $f(X)$ to increase with an increase of x_s from $\bar{x}_s$ to $\bar{\bar{x}}_s$, the variable v_s must be accompanied by $\bar{v}_s \leq \bar{\bar{v}}_s$ because, from (6.11),

$$f(\bar{\bar{X}}) - f(\bar{X}) = -\bar{v}_s(\bar{\bar{x}}_s-\bar{x}_s) + (\bar{\bar{x}}_s-\bar{x}_s)^2 Y'DY$$

and, by interchanging the roles of $\bar{X}$ and $\bar{\bar{X}}$,

$$f(\bar{X}) - f(\bar{\bar{X}}) = -\bar{\bar{v}}_s(\bar{x}_s-\bar{\bar{x}}_s) + (\bar{x}_s-\bar{\bar{x}}_s)^2 Y'DY$$

hence

$$-(\bar{\bar{v}}_s-\bar{v}_s) = 2(\bar{\bar{x}}_s-\bar{x}_s)Y'DY \leq 0,$$

since D is negative semi-definite and $(\bar{\bar{x}}_s-\bar{x}_s) > 0$. Thus we now see that result follows immediately by considering a sequence of feasible solutions to system (6.3) each of which has the value of the variable x_s "slightly" greater than the preceding one.

Q.E.D.

<u>THM. III</u> Let $[X_o, \lambda_o, V_o]$ be complementary basic feasible solution to the system (6.3) and suppose that $v_s^o < 0$ and that one of its basic variables $x_r = x_r^o$, say, drops out of the basic set upon introduction of the variable x_s into the basic set. If the nonbasic complementary variable to x_r, namely v_r, is subsequently increased, then either:

(i) $f(X) = CX + X'DX$ will continue to increase as long as the basic variable v_s remains negative; or

(ii) $f(X) = CX + X'DX$ will stay fixed in which case the basic variable v_s will increase to zero.

PROOF: For reasons of clarity and convenience (specially to avoid messy notation!!) we prove this theorem for the system (6.14) below. <u>It should</u>

be noted that the proof is completely general.

Consider the following system:

$$\begin{bmatrix} A & 0 & 0 \\ 2D & -A' & I_n \end{bmatrix} \begin{bmatrix} X \\ \lambda \\ V \end{bmatrix} = \begin{bmatrix} b \\ -C' \end{bmatrix} \qquad (6.14)$$

where: $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$ $\quad X = [x_1, x_2, x_3, x_4]$

$\lambda = [\lambda_1, \lambda_2, \lambda_3]$

$D = \begin{bmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \\ d_{41} & d_{42} & d_{43} & d_{44} \end{bmatrix}$ $\quad V = [v_1, v_2, v_3, v_4]$

$-C' = [-c_1, -c_2, -c_3, -c_4]$

$I_n = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$ $\quad b = [b_1, b_2, b_3]$

System (6.14) can be compactly represented in the following tableau

x_1	x_2	x_3	x_4	λ_1	λ_2	λ_3	v_1	v_2	v_3	v_4	Constants
a_{11}	a_{12}	a_{13}	a_{14}								b_1
a_{21}	a_{22}	a_{23}	a_{24}								b_2
a_{31}	a_{32}	a_{33}	a_{34}								b_3
d_{11}	d_{12}	d_{13}	d_{14}	a_{11}	a_{21}	a_{31}	1				$-c_1$
d_{21}	d_{22}	d_{23}	d_{24}	a_{12}	a_{22}	a_{32}		1			$-c_2$
d_{31}	d_{32}	d_{33}	d_{34}	a_{13}	a_{23}	a_{33}			1		$-c_3$
d_{41}	d_{42}	d_{43}	d_{44}	a_{14}	a_{24}	a_{34}				1	$-c_4$

(6.15)

Next let the representation of system (6.15) in vector form be

$$P_1x_1+P_2x_2+P_3x_3+P_4x_4+(P_5\lambda_1+P_6\lambda_2+P_7\lambda_3)+\overline{P}_1v_1+\overline{P}_2v_2+\overline{P}_3v_3+\overline{P}_4v_4 = P_o , \quad (6.16)$$

and suppose that we have a basic feasible complementary solution $[X_o,\lambda_o,V_o]$ to (6.16) with basic variables $x_1 = x_1^o \geq 0$, $x_2 = x_2^o \geq 0$, $x_3 = x_3^o \geq 0$, $\lambda_1 = \lambda_1^o, \lambda_2 = \lambda_2^o$, $\lambda_3 = \lambda_3^o$; $v_4 = v_4^o < 0$ (λ_1^o and λ_2^o are of arbitrary sign) and with corresponding basis $B = (P_1,P_2,P_3,P_5,P_6,P_7,\overline{P}_4)$. Now suppose that the value of some basic variable x_r, say x_3 drops out of the basic set, upon introduction of some "non-basic" variable x_s, namely x_4, i.e., suppose that x_4 is being increased and that in the subsequent adjustment of the basic variables, the basic variable x_3 attains the value zero first. In this case, x_4 will become the new basic variable and the vector P_4 will replace the vector P_3 in the basis B. Denote the new basis by $\hat{B}$, then $\hat{B} = (P_1,P_2,P_4,P_5,P_6,P_7,\overline{P}_4)$. Now what we want to show is that if we next increase the complementary variable v_3 (non-basic) of the "dropped" nonbasic variable x_3 (with adjustment of the values of the new basic variables), then:

(a) either the basic variables x_4, v_4 and the objective function $f(X) = CX + X'DX$ will continue to increase as long as v_4 remains negative, or

(b) then $f(X) = CX + X'DX$ will remain unchanged but v_4 will increase to zero.

With this purpose in mind, let the representation of both P_4 and $\overline{P}_3$ in terms of the basis B be:

$$P_4 = P_1\beta_1 + P_2\beta_2 + P_3\beta_3 + (P_5\beta_5+P_6\beta_6+P_7\beta_7) + \overline{P}_4\overline{\beta}_4 \quad (6.17)$$

$$\overline{P}_3 = P_1\omega_1 + P_2\omega_2 + P_3\omega_3 + (P_5\omega_5+P_6\omega_6+P_7\omega_7) + \overline{P}_4\overline{\omega}_4 \quad (6.18)$$

(where the β_i's, the ω_i's, $\overline{\beta}_4$ and $\overline{\omega}_4$ are all scalar coefficients).

Then the scalar coefficient ω_3 which appears in relation (6.18) is non-positive, i.e., $\underline{\omega_3 \leq 0}$. To prove this let $\omega = [\omega_1, \omega_2, \omega_3]$ and eliminate the last row of the system (6.18). Then the resulting system of equations has the following form:

$$\begin{aligned}(a_{11}\ a_{12}\ d_{13})\omega &= \\ (a_{21}\ a_{22}\ a_{23})\omega &= \\ (a_{31}\ a_{32}\ a_{33})\omega &= \end{aligned}\begin{bmatrix}0\\0\\0\end{bmatrix} \qquad (6.19)$$

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}\omega + \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix}\omega_5 + \begin{bmatrix} a_{21} \\ a_{22} \\ a_{33} \end{bmatrix}\omega_6 + \begin{bmatrix} a_{31} \\ a_{32} \\ a_{33} \end{bmatrix}\omega_7 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad (6.20)$$

Hence if we denote the (3×3) square matrix appearing in (6.20) by D_3 and if we premultiply (6.20) by ω' we obtain, on making use of ((6.19),

$$\omega' D_3 \omega = \omega_3 .$$

And thus, since $X'DX$ is a negative semi-definite matrix and since D_3 is a symmetric matrix of D it follows from THM. (XI.3) that

$$\omega_3 = \omega' D_3 \omega \leq 0 \qquad (6.20.a)$$

The reader should satisfy himself that a system of the type of (6.19) and (6.20) is always obtained after deletion of the suitable rows of the system (6.18), by trying out any other possible combination of basic variables and by noting that since we are assuming system (6.3.a) to be non-degenerate the number of columns of A, present in the system (6.19), is never less than m ($m = 3$ in our case).

Next we show that, if $\omega_3 < 0$, then (a) above happens and if $\omega_3 = 0$, then (b) above happens.

Case $\omega_3 < 0$. Let the representation of $\overline{P}_3$ in terms of the new basis $\hat{B} = (P_1\ P_2\ P_4\ P_5\ P_6\ P_7\ \overline{P}_4)$ be

$$\overline{P}_3 = P_1\hat{\omega}_1 + P_2\hat{\omega}_2 + P_4\hat{\omega}_4 + P_5\hat{\omega}_5 + P_6\hat{\omega}_6 + P_7\hat{\omega}_7 + \overline{P}_4\hat{\overline{\omega}}_4 \tag{6.21}$$

and let the new basic solution associated with $\hat{B}$ be

$(\hat{x}_1, \hat{x}_2, \hat{x}_4, \hat{\lambda}_1; \hat{\lambda}_2, \hat{\lambda}_3, \hat{v}_4)$ so that

$$P_o = P_1\hat{x}_1 + P_2\hat{x}_2 + P_4\hat{x}_4 + P_5\hat{\lambda}_1 + P_6\hat{\lambda}_2 + P_7\hat{\lambda}_3 + \overline{P}_4\hat{v}_4 \tag{6.22}$$

Then, substituting (6.17) in (6.21) and equating the scalar coefficients of the resulting expression with the scalar coefficients of expression (6.18) (this is allowed since we are in presence of two basis), we obtain

$$\hat{\omega}_4\beta_3 = \omega_3 \tag{6.23}$$

But

$$B^{-1}P_4 = [\beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_7, \overline{\beta}_4] \qquad \text{from (6.17)}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ v_4 \end{bmatrix} = B^{-1}P_o - B^{-1}.\ P_4x_4 + \overline{P}_1v_1 + \overline{P}_2v_2 + \overline{P}_3v_3 \qquad \text{from (6.5)}$$

and therefore we must have $\beta_3 > 0$, since x_3 dropped out of the basic set, upon introduction of x_4 into the basic set, i.e., since x_3 decreased to zero when the non-basic variable x_4 was increased (keeping v_1, v_2 and v_3 equal to zero of course).

Whence, from (6.23), we conclude that

$$\hat{\omega}_4 < 0$$

since by assumption $\omega_3 < 0$.

And so if $v_3 = \theta > 0$ units of $\overline{P}_3$ are introduced into the system

$$\begin{bmatrix} x_1 \\ x_2 \\ x_4 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ v_4 \end{bmatrix} = B^{-1}P_o - B^{-1}(P_3x_3+\overline{P}_1v_1+\overline{P}_2v_2+\overline{P}_3v_3) \quad \text{with} \quad x_3 = v_1 = v_2 = 0$$

We obtain, from (6.21) and (6.22), that

$$\begin{bmatrix} x_1 \\ x_2 \\ x_4 \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} \hat{x}_1 - \theta\hat{\omega}_1 \\ \hat{x}_2 - \theta\hat{\omega}_2 \\ \hat{x}_4 - \theta\hat{\omega}_4 \\ \hat{\lambda}_1 - \theta\hat{\omega}_5 \\ \hat{\lambda}_2 - \theta\hat{\omega}_6 \\ \hat{\lambda}_3 - \theta\hat{\omega}_7 \\ \hat{v}_4 - \theta\hat{\hat{\omega}}_4 \end{bmatrix} \tag{6.24}$$

Thus $x_4 = x_4 - \theta\omega_4$ will increase when $v_3 = \theta > 0$ is increased since $\omega_4 < 0$. Hence we may adopt the point of view, for the purpose of the proof, that it is the increase in x_4 which is "causing" the increasing in v_3 (instead of the other way around), so that, we are in fact, repeating the situation just considered in the beginning of the proof of increasing x_4 and adjusting the other "basic" variables (remember we assumed that x_3 dropped from the basic set upon introduction of x_4), except here v_3 is in the basic set instead of x_3. And so it follows from THM. II, that an increase in x_4 increases $f(X) = CX + X'XDX$ as long as v_4

remains negative in value in the adjustment of the basic solution by the increase of x_4. This proves statement (i) of our theorem.

Case $\omega_3 = 0$. On the other hand if $\omega_3 = 0$, then relation (6.18) becomes

$$\overline{P}_3 = \sum_{\substack{i=1\\ i\neq 3\\ i\neq 4}}^{7} P_i\omega_i + P_3.0 + \overline{P}_4\overline{\omega}_4 = \sum_{\substack{i=1\\ i\neq 3\\ i\neq 4}}^{7} P_i\omega_i + P_4.0 + \overline{P}_4\overline{\omega}_4 .$$

But from (6.21) we have that

$$\overline{P}_3 = \sum_{\substack{i=1\\ i\neq 3\\ i\neq 4}}^{7} P_i\hat{\omega}_i + P_4\hat{\omega}_4 + \overline{P}_4\hat{\overline{\omega}}_4 .$$

Hence

$$\left.\begin{aligned} \hat{\omega}_1 &= \omega_1 \\ \hat{\omega}_2 &= \omega_2 \\ \hat{\omega}_4 &= \omega_3 = 0, \end{aligned}\right] \qquad (6.24)$$

since $\hat{B} = (P_1, P_2, P_4, P_5, P_6, P_7, \overline{P}_4)$ is a basis

Moreover, from (6.20.a)

$$\omega' D_3 \omega = \omega_3 = 0,$$

and because $X'DX$ is negative semidefinite it follows from THM.(XII.3) that

$$D_3\omega = 0 .$$

This in turn implies $\omega = [\omega_1, \omega_2, \omega_3] = 0$ for if $\omega \neq 0$ the first 3 columns of (6.19) and (6.20) would be linearly dependent, which is impossible because then the square array of coefficients of (6.19) and (6.20), and in turn B, would be singular. Whence, from (6.24), we have that $\hat{\omega}_1 = \hat{\omega}_2 = \hat{\omega}_4 = 0$. Substituting these values in (6.23) we observe that the values of the "basic" variables x_1, x_2 and x_4 remain unchanged with increasing values of $v_3 = \theta > 0$. Furthermore, since v_j and λ_j are not sign restricted, $v_3 = \theta$ can be increased until v_4 drops out of the basis with the value zero (since all x_j are unaffected). This proves statement (ii) of our theorem.

Q.E.D.

From THM. II and III we now see that if we have a complementary basic feasible solution $|X_o, \lambda_o V_o|$, say, to the following quadratic programming Problem.

$$\left.\begin{array}{l} \text{maximise} \quad f(X) = CX + X'DX \\ \text{subject to} \\ \left.\begin{array}{l} AX \qquad\qquad = b \\ 2DX - A'\lambda + V = -C' \end{array}\right] \quad (6.25) \\ X \geq 0 \\ \lambda \quad \text{unrestricted} \\ V \quad \text{unrestricted} \end{array}\right] \quad (Q^2)$$

then: if $V_o \geq 0$, X_o is an optimal solution to the initial Problem (N^2_oII). If on the other hand $v_j^o < 0$ for some j, say, $v_s^o < 0$ then by performing Operations I and II (below) we have the following "set up":

<u>Operation I</u>. Introduce the complementary nonbasic variable x_s of the negative basic variable v_s^o (i.e., increase the nonbasic variable x_s until one of the basic variables present in the basic solution $[X_o, \lambda_o, V_o]$ attains the value zero; note that we do not have to worry about the values of the basic variables λ_i, since these variables are unrestricted). Then:

(a) Either the basic variable v_s drops out of the basic set (i.e., either the basic variable v_s attains the value zero first; note that we do not have to worry if one of the basic variables $v_j \neq v_s$ or λ_j attains the value zero before v_s does so, because these variables are unrestricted), and we obtain a complementary basic feasible solution to Problem (Q^2), say, $[\hat{X},\hat{\lambda},\hat{V}]$ such that

$$f(\hat{X}) > f(X_o). \quad \text{(This follows from THM.II).}$$

(We can then repeat Operation I with the complementary basic feasible solution $[X_o,\lambda_o,V_o]$ replaced by the complementary basic feasible solution $[\hat{X},\hat{\lambda},\hat{V}]$), or

(b) Some basic variable x_j, say, x_r drops out of the basic set (i.e., or some basic variable x_j, say, x_r attains the zero value first) and we obtain a "non-complementary" basic feasible solution to Problem (Q^2) say $[\overline{X},\overline{\lambda},\overline{V}]$ such that

$$f(\overline{X}) > f(X_o). \quad \text{(This follows from THM. II).}$$

(We can now perform Operation II), or

(c) No basic variable x_j, or v_s, attains the value zero no matter how great the increase in x_s. In this case, it follows by THM. II that Problem (Q^2) has an unbounded solution which immediately implies that Problem (N^2.II) also has an unbounded solution since the X-component of any feasible solution to Problem (Q^2) is a feasible solution to Problem (N^2.II).

<u>Operation II</u>. Introduce the complementary non-basic variable v_r of the "just dropped nonbasic" variable x_r, into the basic set (i.e. increase the nonbasic variable v_r until v_s or one of the basic x_j present in the non-complementary basic solution $[X,\lambda,V]$ attains the value zero). Then:

(a) Either v_s drops out of the basic set and we obtain a complementary basic feasible solution to Problem (Q^2), say, $[\hat{X}_1,\hat{\lambda}_1,\hat{V}_1]$, such that

$$f(\hat{X}_1) \geq f(\overline{X}). \quad \text{(This follows from THM.III).}$$

(We can then repeat Operation I with complementary basic feasible solution $[X_o, \lambda_o, V_o]$ replaced by the complementary basic feasible solution $[\hat{X}_1, \hat{\lambda}_1, \hat{V}_1]$), or

(b) Some basic x_j, say, x_{r_1} drops out of the basic set and we obtain a "non-complementary" basic feasible solution to Problem (Q^2) say $[\overline{X}_1, \overline{\lambda}_1, \overline{V}_1]$ such that

$$f(\overline{X}_1) > f(\overline{X}). \quad \text{(This follows from THM. III).}$$

(We can then repeat Operation II with r_1 playing the role of r).

(6.3) Dantzig's Algorithm.

Multiply the constraints $AX = b$, were necessary, by -1, so that the "requirements" vector b becomes non-negative, i.e., so that $b \geq 0$.

Phase I:

STEP I. Test the constraints $AX = b$, $X \geq 0$, for feasibility, using Phase I of the Two-Phase Simplex Method. If the constraints are feasible we obtain a basic feasible solution to the system $AX = b$, say, $X_o = [x_1^o, \ldots, x_m^o, \ldots, x_n^o]$ with basic variables $x_1^o, \ldots, x_m^o$ and with corresponding basis $B = (a_1, \ldots, a_m)$(there is no loss of generality in assuming this particular set of indices for, if the basic variables are, say, $x_{j_1}^o, \ldots, x_{j_m}^o$ then by a suitable reordering of indices we can always obtain them sequenced as $x_1^o, \ldots, x_m^o$; this never needs to be done in practice, it is done here for a more clear presentation of what is happening "behind the scenes").

STEP II. Obtain an initial complementary basic feasible solution $[X, \lambda, V]$ to Problem (Q^2) in the following manner:

(a) Form the system

$$\begin{bmatrix} B' & 0 \\ 0 & I_{n-m} \end{bmatrix} \begin{bmatrix} \lambda \\ V_{n-m} \end{bmatrix} = \begin{bmatrix} b \\ -C' \end{bmatrix} - 2DX_o \qquad (6.26)$$

where $V_{n-m} = [v_{m+1},\ldots,v_n]$ and I_{n-m} = identity matrix of order $(n-m)$, by:

(i) setting $X = X_o$ and $v_1 = v_2 = \ldots = v_m = 0$ in system (6.25)

(ii) deleting the last $n-m$ rows of the matrix A' in system (6.25)

(b) Solve system (6.26), and denote its unique solution by

$$\begin{bmatrix} \lambda \\ V_{n-m} \end{bmatrix} = \begin{bmatrix} \lambda_o \\ V^o_{n-m} \end{bmatrix}$$

Then $[X,\lambda,V] = [X_o,\lambda_o,V_o]$, where $V_o = [0,V^o_{n-m}]$ is a complementary basic feasible solution to the system (6.25). Initiate Phase II.

Phase II.

STEP I. For the non-zero values $v^o_{j_h}$ of the complementary basic feasible solution $[X_o,\lambda_o,V_o]$ determine

$$v^o_s = \text{Min } v^o_{j_h} \qquad h = 1,\ldots,k \ .$$

(Note that in the initiation of Phase II according to our reordering of indices we have $k = n-m$ and $v^o_{j_h} = v^o_{m+h}$ $h = 1,\ldots,n-m$).

If $v^o_s \geq 0$, X_o is an optimal solution to the initial Problem (N^2II).

If $v^o_s < 0$, increase the nonbasic variable x_s. If the basic variable v_s remains negative and the basic variables $x_{j_1},\ldots,x_{j_m}$ (which become $x_1,\ldots,x_m$ in the initiation of Phase II using the reordering of the indices) remain positive for any increase of x_s no matter how great, then Problem (N^2II) has an unbounded solution (see App. VI.3). If one of the basic variables $x_{j_1},\ldots,x_{j_m}$ or v_s attains the value zero, introduce x_s into the basic set. Then:

(a) If v_s drops out of the basic set, repeat Step I with the new complementary basic feasible solution just obtained, say, $[\overline{X}_1, \overline{\lambda}_1, \overline{V}_1]$ in place of $[X_o, \lambda_o, V_o]$.

(b) If some basic x_i drops, say x_{r_1}, repeat Step II with r_1 playing the role of r.

(6.4) <u>Finite Convergence of Dantzig's Algorithm</u>

Suppose that Phase II of Dantzig's Method does not terminate in a finite number of iterations, i.e., suppose that we never reach a complementary basic feasible solution to Problem (Q^2), say $[X_*, \lambda_*, V_*]$ for which $V_* \geq 0$. Then since in at most two "changes" of basis in Phase II we obtain a <u>definite</u> increase in the objective function $f(X) = CX + X'DX$ (compare the steps of Phase II with Operations I and II given after THM. III), it follows that we obtain an infinite sequence of different basic solutions to Problem (Q^2) (not all necessarily complementary note). And this is a contradiction for the number of basic solutions to Problem (Q^2) <u>is finite</u>!! Note also that we have proved a somewhat stronger result than statement (iii) given in section (6.1), namely:– <u>at least one $[X, \lambda, Y]$ satisfying system (6.1) and condition (6.2.b) is a complementary basic feasible solution (instead of simply basic feasible) to system (6.1) provided Problem $(N^2 II)$ has an optimal solution and provided system (6.1) is non-degenerate</u>. This assumption of non-degeneracy implies no loss of generality since any system of equations is treated as if it were not degenerate by the Simplex Method.

(6.5) A Flow of Dantzig's Algorithm in Summary Form.

To save time and space the description of Dantzig's Algorithm given in section (6.3) is already in a Summary Form and so we refer the reader to section (6.3).

Chapter VII

"Beale's Method".

(7.1) Introduction.

Beale's Method is a computational technique developed by E.M.L. Beale for solving a quadratic program in Normal Form II:

$$\left.\begin{array}{l} \text{maximise} \quad f(X) = CX + X'DX \\ \text{subject to} \\ \qquad AX = b \\ \qquad X \geq 0 \\ \text{given that the quadratic form } X'DX \text{ is negative semi-definite.} \end{array}\right] \quad (N^2.II)$$

Here, for the reasons given in Chapter 0 when we discussed the sketch of the present Chapter, we present only a Flow of Beale's Method in Summary Form:

(7.2) A Flow of Beale's Algorithm in Summary Form.

(1) In the system of constraints $AX = b$ choose an $m \times m$ non-singular matrix B and express the m variables associated with the columns of B (called the basic variables) in terms of the remaining $(n-m)$ variables (called the non-basic variables). Then any solution to the constraints $AX = b$ can be written as

$$X_B = B^{-1}b - B^{-1}RX_R \tag{7.1}$$

where: $X_B = [x_{B1}, \ldots, x_{Bm}]$ denotes the variables associated with the columns of A in B.

$X_R = [x_{R1}, \ldots, x_{R,n-m}]$ denotes the variables associated with the columns of A not in B.

R is a matrix which contains the columns of A not in B.

A basic solution to $AX = b$ is obtained by setting in (7.1) $X_R = 0$. The basic solution will be feasible if $B^{-1}b \geq 0$. Now let $B^{-1}b = Y_o = [y_{1o},\dots,y_{mo}]$ and denote the jth column of $B^{-1}R$ by $Y_j = [y_{1j},\dots,y_{mj}]$, then from (7.1) we have that

$$x_{Bi} = y_{io} - \sum_{j=1}^{n-m} y_{ij}x_{Rj} \quad i = 1,\dots,m. \tag{7.2}$$

And so using relation (7.2) we can write the objective function $f(X) = CX + X'DX$ of Problem (N^2.II) in the form

$$f_R(X) = \alpha + \beta X_R + X_R'GX_R \tag{7.3}$$

where α is a scalar, β a constant row vector and G an $(n-m)\times(n-m)$ constant matrix.

(2) Suppose now that $X_B = Y_o$ is a basic feasible solution to the system of constraints $AX = b$, and that $\frac{\partial}{\partial x_{Rj}} f_R(X) > 0$ for some j, say $j = k$. Then a small increase in x_{Rk}, with the other non-basic variables (x_{Rj} $j \neq k$) held equal to zero, will increase $f_R(X)$. Hence it is profitable to go on increasing x_{Rk} until either (a) some x_{Bi}, say x_{Br}, becomes zero and decreases to negative values or (b) $\frac{\partial}{\partial x_{Rk}} f_R$ becomes zero and decreases to negative values. If (a) happens the set ϕ nonbasic variables is changed be replacing x_{Bk} by x_{Br}, and $f(X)$ is expressed in terms of the "new" nonbasic variables and Step (2) is repeated. In (b), as $\frac{\partial}{\partial x_{Rk}} f_R(X)$ is a linear function of the x_{Rj} (since $f_R(X)$ is a quadratic function), we introduce

$$\mu_k = \frac{1}{2}\frac{\partial}{\partial x_{Rk}} f_R(X)$$

as a nonbasic variable. This μ_k is similar to any other nonbasic variable, except that μ_k may assume both positive and negative values; hence it is called a <u>free variable</u>, as distinguished from the other nonbasic variables which cannot be negative. The objective function $f(X)$ is again expressed in terms of the "new" set of nonbasic variables and Step (2) is repeated. The method converges in a finite of steps and after a finite number of iterations we shall obtain

$$\frac{\partial}{\partial x_{Rj}} < 0 \qquad j = 1,\ldots,n-m$$

This implies that we are in presence of a local maximum for the objective function $f(X)$. But since $f(X)$ is a concave function in E_n this implies that we are in presence of a global maximum for $f(X)$ (see THM.VI*.1). A most lucid and exhaustive article on Beale's Method can be found in (11). (3), (15) and (16) also provide good descriptions.

APPENDIX I

(1) Taylor's Theorem states that: If $f(X) \in C'$ over E_n then for any two points X_1 and X_2 in E_n there exists a θ, $0 \leq \theta \leq 1$, such that

$$f(X_2) = f(X_1) + \nabla f(\theta X_1 + (1-\theta)X_2)(X_1-X_2)$$

(see (15) Chapter II for a proof of this theorem).

(2) The empty set is, by convention, convex and any set with only one element is clearly convex.

(3) A detailed proof of this statement can be found in (15) Chapter III, section (3.12).

APPENDIX II

(1) If $R_X = \{X | AX \leq b, X \geq 0\}$, then R_X is a closed convex set. A detailed proof of this statement can be found in (14), Chapter II, sections (2.20) and (2.21).

(2) Since in THM. I we do not use the fact that $f \in C'$ and $g_i(X) \in C'$ for $i = 1,\ldots,m$, THM. I applies equally well to any type of Concave Program, i.e., to any problem of the type

$$\text{maximise} \quad f(X)$$

subject to

$$g_i(X) \leq 0 \qquad i = 1,\ldots,m$$

$$X \geq 0$$

given that $f(X)$ is a concave function in E_n.

APPENDIX III

(1) Hildreth's Method (see (16)) can only be applied to quadratic programs where the quadratic objective function $f(X) = CX + X'DX$ is strictly concave (i.e., where D is strictly negative definite). The Method of Theil and Van de Panne (see (16)) is another computational technique which is limited to quadratic programs where the objective function is strictly concave.

(2) Here we prove the following theorem:

Let $f(X) = CX + X'DX$ (C a constant vector, D a constant symmetric matrix, $X \in E_n$). Then if $f(X)$ is bounded above on a polyhedral convex set R, it assumes its supremum on R.

For the proof of this theorem we need some background theory:

(I) Polyhedral Convex Sets.

(a) A polyhedral convex set is the intersection of a finite number of closed half-spaces, and so the set

$$\{X \mid AX \leq b, \quad X \geq 0\}$$

is a polyhedral convex set.

Notes:

(1) A polyhedral convex set is closed (see: A.J. Goldman : Resolution and Separation Theorems for Polyhedral Convex Sets (page 41) in Linear Inequalities and Related Systems (edited by Kuhn and Tucker, Princeton 1956). This reference will be referred to as (G).

(2) The intersection of two polyhedral convex sets is again such a set (easily seen from the definition).

(3) Once the above theorem is proved THM. VI.3 and THM. VII.3 follow immediately since $\mathbb{K}_S$ and $\mathbb{K}_N$ are polyhedral convex sets.

THM: Any polyhedral convex set R can be written as the sum of a bounded convex polyhedron S and a cone Q

i.e. $R = \{s+q | s \in S, q \in Q\} = S + Q$

Proof: (See (G) THM. I page 44).

Note — it is easily seen that S and Q must be closed.

(c) Let Q be a cone, and let T be the intersection of Q with the unit sphere (not the unit ball, but just its boundary). Then

$$Q = \{\mu t | t \in T, \mu \geq 0\}.$$

Proof: Since Q is a cone,

$$q \in Q \Longleftrightarrow \mu q \in Q, \mu \geq 0. \qquad \text{(i)}$$

Then if $q \neq 0$, $q = |q| \cdot (q/|q|)$. But $|(q/|q|)| = 1$ and therefore $q/|q| \in$ unit sphere. It also belongs to Q by (i) and hence $q/|q| \in T$, and $q = \mu^* = |q| > 0$.

Now if $q = 0$, $q = 0 \cdot t$ for any $t \in T$ and therefore

$$Q \subset \{\mu t \ t \in T, \mu \geq 0\} \qquad \text{(ii)}$$

Also, if $t \in T$, then $t \in Q$ and $\mu t \in Q$ $(\mu \geq 0)$ by (i) and hence

$$\{\mu t | t \in T, \mu \geq 0\} \subset Q \qquad \text{(iii)}$$

Result follows from (ii) and (iii).

(d) From (c) we see that the conclusion of (b) may be changed to read

$$R = \{S + \mu t | s \in S, \mu \geq 0, t \in T| \quad \text{where}$$

$$T = Q \cap \{\text{unit sphere}\}.$$

Note: it is clear that $S \subset R$ (case $\mu = 0$).

(e) If a half-line of the form $L : r + \mu t$, $\mu \geq 0$, belongs to R, then so does the half-line $r^* + \mu t$, $\mu \geq 0$, for any $r^* \in R$.

Proof: Let $r^* \in R$. Since R is convex, any line segment M, joining r^* to a point on L, belongs to R. But the closure of these segments contains the half-line $r^* + \mu t$, $\mu \geq 0$. Since R is closed it follows that this half-line is in R.

Notes:

(1) This works also for half-lines of the form

$$L_1: r + \mu t \ , \ \mu \leq 0.$$

(2) Thus, if the set R contains a line of the form $r + \mu t$, all μ (μ arbitrary, not necessarily ≥ 0), then $r^* + \mu t$, all μ, belongs to R for any $r^* \in R$. (By above result, $r^* + \mu t$, $\mu \geq 0$, belongs to R, and by note (1), $r^* + \mu t$, $\mu < 0$ belongs to R).

(3) Thus for a fixed t_o, either $r + \mu t_o$, all μ, belongs to R <u>for each</u> $r \in R$, or for each $r \in R$, there exists values μ^* of μ such that $r + \mu^* t_o \notin R$.

(4) Since for each $t \in T$, we have $s + \mu t \in R$, $\mu \geq 0$, it follows that $r + \mu t \in R$, $\mu \geq 0$, all r. Therefore if $t_o \in T$, and $r + \mu t_o \in R$, all $\mu \geq 0$ does <u>not</u> hold, and then the values of μ for which $r + \mu t_o \notin R$, must satisfy $\mu < 0$.

<u>(II) We are now ready to give a proof of our theorem:</u>

Proof: By induction on the dimension n:

(a) $n = 1$: $X \in E_1$ $f(X) = c_1 x_1 + d_{11} x_1^2$; $c_1, d_{11} \in E_1$.

A polyhedral convex set is a closed interval or a half-line. Clearly $f(X)$ assumes its maximum on any of these sets, since $f(X)$ bounded above implies that maximum is assumed (if set is bounded) or implies $d_{11} < 0$ or $c_1 = 0$, $d_{11} = 0$ (if set is unbounded).

(b) Assume true for $n = k$. R.T.P. for $n = k + 1$.

Suppose that $f(X)$ is bounded above on the polyhedral convex set R in E_{k+1}. Then $R = \{s + \mu t \mid s \in S, t \in T, \mu \geq 0\}$ as in I(d). S is bounded above and T is bounded above. For any r, t, μ:

$$f(r+\mu t) = C(r+\mu t) + (r+\mu t)'D(r+\mu t)$$

$$= Cr + r'Dr + \mu(C+2r'D)t + \mu^2(t'Dt)$$

$$\therefore f(r+\mu t) = f(r) + \mu(C+2r'D)t + \mu^2(t'Dt) \qquad \text{(iv)}$$

In particular, for $s \in S$ and $t \in T$, $\mu \geq 0$:

$$f(s+\mu t) - f(s) = \mu(C+2s'D)t + \mu^2(t'Dt) \qquad \text{(v)}$$

Since $f(X)$ is bounded above, L.H.S. is bounded above (on R) $\therefore$ R.H.S. bounded above on R ($\mu \geq 0$), therefore $t'Dt \leq 0$. We have two cases: $t'Dt < 0$ all ($t \in T$) or there exists $t_o \in T$ such that $t_o'Dt_o = 0$.

<u>Case (1) — Assume $t'Dt < 0$ (all $t \in T$)</u> (vi)

Since T is compact (closed and bounded), it follows that $\sup_{t \in T} (t'Dt)$ is attained on T ($t'Dt$ is a continuous function).

$\therefore \sup_{t \in T} t'Dt = t_1'Dt_1$ for some $t_1 \in T$. Therefore $\sup t'Dt < 0$ by (vi) and so $t'Dt < -\delta$, for all $t \in T$, some $\delta > 0$. Since S and T are bounded above there exists a real number d such that

$$(C+2s'D)t < d \quad \text{for all } s \in S, \ t \in T.$$

For fixed s and t the maximum of R.H.S. of (v) is taken at

$$\mu = -\frac{(C+2s'D)}{t'Dt} \leq \frac{d}{\delta}$$

Therefore for each (fixed) s and t, maximum is taken at a point $s + \mu t$ with $\mu \leq \frac{d}{\delta}$. Thus

$$\sup_{r \in R} f(r) \leq \sup_{r \in S + \frac{d}{\delta}T} f(r).$$

Since $S + \frac{d}{\delta} T \subset R$, we have that

$$\sup_{r \in S + \frac{d}{\delta}T} \leq \sup_{r \in R}$$

and therefore

$$\sup_{r \in S + \frac{d}{\delta}T} f(r) = \sup_{r \in R} f(r) .$$

Since sup on L.H.S. is attained at a point in $S + \frac{d}{\delta}T$, i.e., at a point in R it follows that $\sup_{r \in R} f(r)$ is attained at a point in R.

<u>Case (2) — Assume there exists t_o such that $t_o' D t_o = 0$</u>

We have two possibilities:

(2.a) Either for all r, $r + \mu t_o \in R$, <u>all</u> μ (not only $\mu \geq 0$) (vii)

(2.b) Or for each r, $r + \mu t_o \in R$ for some $\mu < 0$ (viii)

(this follows from I(e) notes (3) and (4)).

<u>Case (2.a)</u> - Assume (vii) holds:

Now $f(r + \mu t_o) = f(r) + \mu(C+2r'D)t_o$. This must be bounded above for <u>all</u> μ and therefore $(C+2r'D)t_o = 0$.

Hence $f(r+\mu t_o) = f(r)$ (all $r \in R$ and all μ) (ix)

Now any vector $X \in E_{k+1}$ can be written in the form

$$X = U_x + \gamma_x t_o \quad \text{for some} \quad \gamma_x \in E_1 ,$$

where U_x belongs to the k-dimensional subspace, $\mathbb{U}$ say, which is orthogonal to the subspace $\{\mu t_o | \mu \in E_1\}$ generated by t_o. For $r \in R$, $r = U_r + \gamma_r t_o$ and therefore

$$U_r = r - \gamma_R t_o \in R \quad \text{by (vii)}$$

Thus $f(r) = f(U_R + \gamma_r t_o) = f(U_r)$ by (ix) and so

$$\sup_{r \in R} f(r) = \sup_{U_r \in \mathbb{U}_R} f(U_r)$$

where

$\mathbb{U}_R = \{U_r |$ there exists $r \in R$ and γ_r such that $r = U_r + \gamma_r t_o\}$. But $\mathbb{U}_R$ is at most k-dimensional, and, a polyhedral convex set ($\mathbb{U}_R$ is the intersection of an hyperplane $\mathbb{U}$ with R. Both $\mathbb{U}_R$ and R are convex polyhderal sets in E_{k+1}, and it is clear (I(a) note 2) so is their intersection. It is also clear that a convex polyhedral set in E_{k+1} which contains a k-dimensional subspace is a convex polyhedral set in E_k). Therefore by the inductive hypothesis the supremum is assumed at U_r^* say. We showed above that $U_R^* \in R$ and therefore the supremum of $f(X)$ is assumed on R.

Case (2.b) — Assume (viii) holds.

Since $f(r+\mu t_o) = f(r) + \mu(C+2r'D)t_o$, and this is bounded above for $\mu \geq 0$, we must have $(C+2r'D)t_o \leq 0$. Therefore

$$f(r+\mu t_o) \geq f(r) \qquad \text{for } \mu < 0, \tag{x}$$

By (viii), for some μ_r, $b_r = r + \mu_r t_o$ must belong to the boundary of R. By (x), $f(b_r) \geq f(r)$ therefore

$$\underset{U \in \text{Boundary of } R}{\text{supremum}} f(U) \geq \underset{r \in R}{\sup} f(r)$$

But Boundary of $R \subset R$ and hence

$$\underset{U \in \text{Boundary of } R}{\text{supremum}} f(U) = \underset{r \in R}{\sup} f(r)$$

But the boundary of a (k+1) dimensional polyhedral convex set is again a polyhedral convex set (it is a union of the "faces" of the polyhedral convex set, which are each polyhedral convex sets). Thus by inductive hypothesis

$$\underset{U \in \text{Boundary of } R}{\text{supremum}} \text{ is attained on Boundary of } R \subset R$$

Q.E.D.

B I B L I O G R A P H Y

PAPERS:

(1) BARANKIN, E., and K. DORFMAN, "A Method for Quadratic Programming", Econometrica, Volume 24, 1956.

(2) BEALE, E.M.L., "On Minimizing a Convex Function Subject to Linear Inequalities", Journal of the Royal Statistical Society (B), Volume 17, 1955, pp. 173-184.

(3) BEALE, E.M.L., "On Quadratic Programming", Naval Research Logistics Quarterly, Volume 6, 1959, pp. 227-243.

(4) DANTZIG, G.B., "General Convex Objective Forms", in K.J. Arrow, S. Karlin and P. Suppes "Mathematical Methods in the Social Sciences", Stanford University Press, Stanford, California, 1960, pp. 151-158.

(5) DANTZIG, G.B., "Quadratic Programming : A Variant of Wolfe-Markowitz Algorithm ", Operations Research Center, University of California, Berkeley, R.R.2 (April 1961).

(6) FRANK, M., and P. WOLFE "An Algorithm for Quadratic Programming", Naval Research Logistics Quarterly, Volume 3, 1956, pp. 95-110.

(7) KUHN, H.W., and A.W. TUCKER, "Nonlinear Programming", Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, J. Newman, editor, pp. 481-492. (Berkeley and Los Angeles : University of California Press, 1951).

(8) MARKOWITZ, H., "The Optimization of a Quadratic Function Subject to Linear Constraints", Naval Research Logistics Quarterly, Volume 3, 1956, pp. 111-133.

(9) WHINSTON, A., and C. VAN DE PANNE, "Simplicial Methods for Quadratic Programming", Naval Research Logistics Quarterly, Volume 12, 1965 pp.273-294.

(10) WOLFE, P., "The Simplex Method for Quadratic Programming", Econometrica, Volume 27, 1959, pp. 382-398.

BOOKS:

(11) ABADIE, J., "Nonlinear Programming", North-Holland Publishing Company, Amsterdam, 1967.

(12) DANTZIG, G.B., "Linear Programming and Extensions", Princeton University Press, Princeton, New Jersey, 1963.

(13) DUFFIN, J.F., E.L. PETERSON and C.M. ZENER, "Geometric Programming", John Wiley & Sons, New York, 1967.

APPENDIX IV

(1) An initial basic feasible vector for the linear programming problem maximise $z = CX$ subject to $AX = b \geq 0$, $X \geq 0$ is found by applying the simplex method to the extended linear programming problem maximise $z^* = -\sum_{i=1}^{m} y_i$ subject to $AX + I_n Y = b$; $X \geq 0$; $Y \geq 0$; $X = 0$ and $Y = b \geq 0$ being an initial basic feasible vector for this problem. If $\text{Max } z^* < 0$, then the original problem is not feasible; otherwise, the optimal solution $[X_*, Y_* = 0]$ which is obtained for the extended problem is a basic feasible solution to the original problem. See (14), Chapter V.

(2) The reader is referred to (14) Chapter VI for an exhaustive discussion of degeneracy in the Simplex Algorithm.

(3) A linear programming problem falls into one of three mutually exclusive and collectively exhaustive categories:

(i) It has an optimal solution.

(ii) It has an unbounded solution.

(iii) It has no feasible solution.

(See remark after THM. (XIV.3)).

(4) Let X_B be a basic feasible solution with corresponding basis B and corresponding price vector C_B, to the problem maximise $z = CX$, subject to $AX = b$, $X \geq 0$, then

$$z_j = C_B B^{-1} a_j \text{ for } j = 1,\dots,n,$$

where a_j is a column of the matrix A is defined to be the marginal coefficient of the above problem (corresponding, of course, to the basic feasible solution X_B).

(5) For a proof of the Fundamental Theorem of Duality refer to (14) Chapter VIII section (8.3).

(6) For a proof of the Complementary Slackness Property refer to (14) Chapter VIII section (8.5).

(7) For a detailed description of the Long Form we refer the reader to (10).

APPENDIX V

(1) Please refer to App. (IV.1)

(2) In the likely event that the constraints of Problem (A^*) are degenerate "$\leq$" may occur here for a while, but not for long. Dantzig's Method for handling degeneracy (12) is exceptionally easy to use here.

(3) If the linear form $z = CX$ is bounded above in the feasible domain $\{X|AX \leq b, X \geq 0\}$ then it attains its supremum there (see remark (ii) after THM. VI. 3).

(4) A convex polyhedron is a closed and bounded set. For a proof of this statement the reader is referred to (14) Chapter II, sections (2.22) and (2.23).

(5) The proof of THM. (IX.3) follows since the maximum of $z = -(V'X+\lambda'Y) = -\frac{1}{2}\bar{Z}Z$ over $\mathbb{K}_A$ is equal to zero.

APPENDIX VI

(1) There is no loss of generality in assuming that system (6.3) is non-degenerate because the solutions of system (6.3) are continuous functions of appropriately chosen samll perturbations in the "constants" of the system. Thus we can keep the changes in the solutions arbitrarily small by a sufficiently small perturbation. A detailed proof of this statement can be found in "Management Models and Industrial Applications of Linear Programming" by A. Charnes and W.W. Cooper (Volume II, pp 682-687; John Wiley & Sons, New York, 1961).

(2) Since z_{Br} goes to zero upon increasing z_{Rk}, it follows that $y_{rk}>0$ ($\neq 0$) and hence we obtain a "new" basic solution upon replacing z_{Br} by z_{Bk} in the basis (i.e. in the basic set).